Summer 8-18-2010

Essays on the Effects of Early Childhood Malnutrition, Family Preferences and Personal Choices on Child Health and Schooling

Solomon T. Tesfu
Georgia State University

PERMISSION TO BORROW

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University Library must be used only in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this dissertation is:

Solomon T. Tesfu
214 Barry St.
Decatur, GA 30030

The director of this dissertation is:

Shiferaw Gurmu
Department of Economics
Andrew Young School of Policy Studies
Georgia State University
P.O. Box 3992
Atlanta, Georgia 30302-3992

Users of this dissertation not regularly enrolled as students at Georgia State University are required to attest acceptance of the preceding stipulations by signing below. Libraries borrowing this dissertation for the use of their patrons are required to see that each user records here the information requested.

Name of User	Address	Date	Type of use (Examination only or copying)

ESSAYS ON THE EFFECTS OF EARLY CHILDHOOD MALNUTRITION, FAMILY PREFERENCES AND PERSONAL CHOICES ON CHILD HEALTH AND SCHOOLING

BY

SOLOMON TESFAY TESFU

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY
2010

ACCEPTANCE

This dissertation was prepared under the direction of the candidate's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair: Shiferaw Gurmu

Committee: José J. Canals-Cerda
Barry T. Hirsch
Erdal Tekin

Electronic Version Approved:
Mary Beth Walker, Dean
Andrew Young School of Policy Studies
Georgia State University
August 2010

ACKNOWLEDGEMENTS

I extend my sincere appreciation and gratitude to all who have assisted me while working on this dissertation. I owe more than I can say to my dissertation committee chair, Dr. Shiferaw Gurmu, for he has been involved since the early stages and provided critical inputs at every step of the way. I can't thank him enough not only for his intellectual guidance of high caliber in the process of writing this dissertation but also for his unreserved support and encouragement throughout my stay in the graduate program. Dr. Gurmu and his family have treated me like a family member for the last five years and I'll never forget their generosity.

No words of gratitude will do justice to the enormous contributions from my dissertation committee member, Dr. Barry Hirsch. This dissertation has certainly become a much better product due to his incredible wisdom, depth, dedication, keen interest, and attention to detail out of which I have learned a lot. I'm sincerely grateful for his extraordinary kindness and friendly treatment that made me feel good about what I was doing every time I met him.

I have also benefited from the suggestions and insights of the other members of my dissertation committee, Dr. Erdal Tekin and Dr. Jose Canals-Cerda. I would like to thank Dr. Tekin, for his suggestions and challenges especially at the early stages of the dissertation for they have shaped my thinking and helped improve the final product. I gratefully acknowledge Dr. Canals-Cerda's kindness in accepting to serve as external member of my dissertation committee even though we have never seen each other physically. His comments and suggestions both on the proposal and the draft of the dissertation were quite helpful.

I have also gained a lot from Dr. Paula Stephan, Dr. Ragan Petrie, and Dr. Inas Rashad who have kindly served as readers of my dissertation proposal. Their suggestions and comments on the proposal have helped me improve the quality of all the three essays and I'm grateful to all of them. I would also like to thankfully acknowledge Dr. Petra Todd of the University of Pennsylvania and Dr. Umut Ozek of the Urban Institute for their suggestions that have helped improve Essay I.

My special thanks, however, go to my family whose love and support has been the only constant in my life. My mom, Abeba Woldeyohannis, had to endure extraordinary challenges to raise eight of us as a widow since I was two years old and no words in this world will be enough to express the love and admiration I have for her. Her deep faith in God and relentless prayers are what gives me strength when I'm down and I'm eternally grateful for everything she did and still does for me.

My older brother, Seyoum Tesfay, had to put his future on hold for eight years to help the rest of us have a shot at a bright future. His guidance and inspiration at my early childhood is what still keeps me going and I'm deeply grateful for all the sacrifices he made for me and my other siblings.

My sister, Saba Tesfay, has been the pillar of our family for over two decades. Without her limitless financial support, generosity, love and encouragement I would have never been where I'm today and she will always have a special place in my heart. The best times I have had over the last five years have always been the weeks I spent with Saba and her lovely kids, Memona and Biruh. I'm very proud to have her as a sister.

We were both rivals and friends with my younger brother, Mekonnen Tesfay, with whom I have started school in the same class. Intellectual competition with him made me smarter and I'm very grateful for everything we did together. I can never forget the kindness of my sisters Abrehet, Fissiha, and Nigisti, and the honesty, kindness and hard work of my brother Fissihaye. They have made our family more complete and I'm proud of all of them.

My 'thank you' notes will be incomplete without extending my heartfelt gratitude to Bel Haile particularly for being around at one of the most critical times in my life. I owe a lot to her.

Finally, I would like to thank the Department of Economics and the Andrew Young School of Policy Studies for generously funding my graduate education. I'm especially grateful to the faculty and staff of the Department of Economics for all their support and kindness throughout my stay in the graduate school.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ESSAYS ON THE EFFECTS OF EARLY CHILDHOOD SHOCKS, FAMILY PREFERENCES AND PERSONAL CHOICES ON CHILD HEALTH AND SCHOOLING

ABSTRACT

SOLOMON TESFAY TESFU

AUGUST 2010

Committee Chair: Dr. Shiferaw Gurmu

Major Department: Economics

This dissertation consists of three essays investigating the role of early life events, family environment and personal choices in shaping a child's chances for human capital accumulation. The first essay examines how physical stature of a child measured in terms of age standardized height influences his/her selection for family labor activities vs. schooling in rural Ethiopia using malnutrition caused by exposure to significant weather shocks in early childhood as sources of identification for the child's physical stature. We find no evidence that better physical stature of the child leads to his/her positive selection for fulltime child labor activities. On the other hand we found reasonably strong and consistent evidence that physically more robust children are more likely to combine child labor and schooling than physically weaker children. The findings indicate that, although better early childhood nutrition leads to higher chances of attending school, it may also put the child at additional pressure to participate in family labor activities which may be reflected in poor performance in schooling.

The second essay empirically investigates whether the quantity deficit in the children of the mother's preferred gender is compensated through their favorable treatment in terms of investment in schooling and nutrition (referred to as compensating hypothesis) and to what extent the mother uses her bargaining power in the family to influence this process. We use data from siblings and twins in two rounds of the demographic and health surveys of Ethiopia with robustness checks using a similar but larger data set from India. We find the mother's bargaining power working in the opposite direction to that of the compensating hypothesis in the case of child schooling and having no substantive role in the case of child nutritional health. Our findings for child schooling make intuitive sense in the context of Basu's (2006) hypothesis which implies that mother's empowerment could turn out to be unfavorable to a child's attendance of schooling in the circumstances where the child is needed to help out with family activities.

In the third essay we use date from the 1997 cohort of the National Longitudinal Survey of the Youth (NLSY97) to examine the extent to which high school completion (and to a limited extent college enrollment) are influenced by the choice teenagers make as to when to start dating and/or engage in sex, how many dating and/or sex partners to maintain, and how frequently to engage in sexual and/or dating activities. We use indicators of parental and peer religiosity as instruments for teenager's involvement in sex and dating activities. While our results for teenage dating are generally weaker than those for teenage sex, the overall pattern of our estimates suggests that teenage sex and dating could have significant effects not only on high school completion but also the subsequent enrollment in a college.

INTRODUCTION AND SUMMARY

The amount of skill, knowledge and capability embodied in labor (human capital) is a key not only for individual success in the labor market but also for the progress of a society at large. A healthy and educated person with good work ethic and experience not only earns a higher income for him/herself in the labor market but may also contribute to the advancement of society. And the foundation for successful accumulation of human capital is laid early in life beginning with good nutrition and family-value environment followed by timely access to and continuous attendance of acceptable quality schooling. Consequently, differences in human capital accumulation start to arise early in life because of a number of factors that could broadly be classified into three categories. First, factors beyond the control of both the family and the child such as availability of schooling and health facilities, government policies, and differential exposure to market and environmental shocks could explain some of the observed differences across individuals in early human capital accumulation. Second, differences in family resources, endowments and preferences could be responsible for the differences in early human capital accumulation. Third, individual differences in abilities, preferences and behavioral choices made early in life could explain the observed differences in early human capital accumulation. Various components of these factors have been subject to wide-ranging theoretical and empirical research both in economics and other social sciences. A general observation that could be made from this large literature is that human capital accumulation is a complex process influenced by many interdependent variables that may vary from one place to another and from one context to another (see Heckman 2008).

This dissertation contributes towards better understanding the reasons behind the observed differences in human capital across individuals and groups by picking up one new question under each of the three broad categories listed above. The first essay examines how malnutrition caused by exposure to significant weather shocks (i.e., a massive drought and rainfall fluctuations) early in a child's life influence the subsequent participation of the child in schooling and child labor activities. Most of the empirical studies on the effect of early childhood malnutrition find negative relationship between schooling and cumulative long-term nutrition status and suggest that it is because parents are less likely to send a physically weaker child to school compared to physically robust child. However, in the communities where children are needed for family/child labor, parents may see physical strength as an important factor for the effectiveness of the child in non-schooling activities as well and may actually be tempted to keep some of the physically stronger children for such activities as demonstrated by our theoretical model. To test the empirical validity of this hypothesis, bivariate probit and multinomial logit models are estimated using data from various rounds of Ethiopian Rural Household Survey (ERHS).

We find no conclusive evidence that better physical stature of the child leads to his/her positive selection for fulltime child labor activities. On the other hand we found strong and consistent evidence that physically robust children are more likely to combine child labor and schooling than physically weaker children. The results are consistent across two different identification strategies applied to data from two different cohorts of children. The findings indicate that, although better early childhood nutrition leads to higher chances of attending school, it may also put the child at additional pressure to participate in family labor activities and this may be reflected in poor performance in schooling. Therefore, policies that try to promote schooling through nutrition support programs could be more successful if they are accompanied by programs that could mitigate the forces that push families to resort to child labor.

The second essay relates to the general issue of the role of family resources and preferences and specifically asks whether the explicitly revealed preferences of the mother for the gender mix of her children coupled with her bargaining power in the family influence the distribution of investment in human capital across her children. The essay examines the important question of whether the mother uses her bargaining power in the family to influence the distribution of child quality in the direction of her gender ratio preferences when the actual gender ratio of children deviates from her preferred ratio. For example, if the mother prefers to have three boys and three girls but ends up with one girl and 5 boys, does she try to compensate for the deficit in the quantity of girls by using her influence on household resource allocation to more heavily invest in her give daughter's quality (human capital)? Essay II attempts to shed some light on this question using data on twins and siblings from two rounds of the demographic and health survey of Ethiopia with some robustness checks using a similar but larger data set from India. Specifically, the role of the interaction between the mother's gender ratio gap and her bargaining power in the distribution of child health and schooling are examined.

Our results provide no conclusive support to the hypothesis that the mother tries to compensate for the deficit in the quantity of children of her preferred gender through preferential treatment in their schooling and nutritional health although there are qualitative indications in some of our results that this could be the case. While our expectation was that higher mother's power will reinforce the effect of mother's gender ratio gap on the distribution of child schooling and nutritional health in the direction implied by the compensating hypothesis, we rather find it working in the opposite direction in the case of child schooling having no substantive role in the case of child's nutritional health. Our results for schooling make intuitive sense in the context of Basu's (2006) hypothesis, which implies that mother's power could be unfavorable to a child's attendance of schooling in the contexts where the child is needed to help out with family activities. Although we do not want to stretch the policy implications of these results too far before they are confirmed in other contexts with different data sets and estimation techniques, they seem to indicate that women's empowerment programs should be supplemented with other policies that mitigate their needs for child labor (like income support programs) in order to enhance child schooling.

Finally, the third essay investigates the effect of some choices made by the children themselves on their success at school. Specifically, the essay examines how children's schooling outcomes are affected by their involvement in dating and sexual activities as teenagers. While a number of empirical studies investigated the effects of other choices made by children such as drug use, alcohol drinking and smoking, and involvement in crime and gang membership, there is relatively little research on the effects of teenage dating and sex on schooling outcomes, apart from the literature on how teenage pregnancy and child bearing affect schooling. Even when teenage sex and dating do not result in pregnancy and child birth, there is reason to believe that intense engagement of a teenager in such activities could affect his/her schooling outcome because of the opportunity cost of the time spent in such activities that otherwise may have been used for studying and perhaps because such activities may affect the focus and attention of the child.

We use data from the 1997 cohort of the National Longitudinal Survey of the Youth (NLSY97) to examine the extent to which high school completion (and to a limited extent college enrollment) are influenced by the choice teenagers make as to when to start dating and/or sex, how many dating and/or sex partners to maintain, and how frequently to engage in sexual and/or dating activities. We use indicators of parental and peer religiosity as instruments for teenager's involvement in sex and dating activities. While our results for teenage dating are generally weaker than those for teenage sex, partly because of poor performance of our instruments for the former, the overall pattern of our estimates suggests that teenage sex and dating could have significant effects not only on high school completion but also the subsequent enrollment in a college. We observe only small changes in the estimates when we control for teenage pregnancy and child birth implying that the indicators of teenage sex and dating are not just capturing the effects of the former. The fact that parental and peer religiosity appears to constrain the teenagers' involvement in sex and dating (the effect being stronger on teenage sex) and the fact that the latter influence schooling outcomes implies that religious morality could be one (but not the only) way to influence schooling outcomes by imposing more discipline on the teenagers in terms of delaying initiation of sex and dating, limiting sex/dating partners and frequency of sex/dating. This poses an important policy dilemma (including some Constitutional issues) regarding the extent to which religious morality ought to be promoted in order to improve schooling outcomes.

ESSAY I: THE EFFECT OF EARLY CHILDHOOD MALNUTRITION ON CHILD LABOR AND SCHOOLING: EVIDENCE FROM ETHIOPIA

Introduction

Unlike the developed economies where short-term fluctuations in household income and living standards are largely associated with the conditions in the labor market and business cycles, temporary changes in livelihoods of rural communities in the least developed economies are often caused by changes in weather conditions. In such communities, large and unexpected changes in weather conditions can sometimes have a devastating impact on income, consumption, assets, health and survival of households and their members. Drought, flooding, hailstorms, cyclonic storms, and frost are some of the weather related shocks that frequently affect the livelihoods of rural communities in developing countries. A large number of studies have investigated the various impacts of such shocks and how households try to cope with their effects. The overall picture that emerges from the multitude of empirical studies is that the ultimate impact of a shock on the well-being of a household and its members depends on a number of household and community-specific characteristics such as liquidity constraints, wealth status, and the nature and capabilities of social support networks to which households belong (see Townsend 1995; Murdoch 1999; Carter and Maluccio 2003).

One important indicator of the capability of households to absorb the effects of a shock is whether the nutritional status of its members, as reflected in anthropometric health measures, substantially deteriorates as a result of the shock. While some evidence shows that adults may lose some body mass (Dercon and Krishnan 2000) and may sustain some long-term deficiencies in their health and fitness as a consequence of shocks, there is a general agreement in the literature that it is children in the early years of life who sustain the greatest long-term damage in their physical stature and possibly cognitive ability (Dasgupta 1997). The majority of empirical studies show that children in their first 3 years of life at the time of the shock are particularly vulnerable. This is not surprising given that this is a period when children are growing fast and have high nutritional requirements per unit of body mass (Martorell et al. 1995; Martorell 1999; Hoddinott and Kinsey 2001). Another reason for the high nutrition requirements for young children is their vulnerability to diseases because of immature immune systems and the inability to make their needs known.

Some studies have examined the extent to which exposure to a shock at this early age affects the human capital outcomes of the person later in life. While some evidence from the United States shows that reversal of the effects of early malnutrition is possible if there are dramatic favorable changes in the environment for the child at the appropriate time (Golden 1994), studies from developing countries (e.g., Alderman, Hoddinott and Kinsey 2006) show that victims of severe shocks in early childhood often sustain long-lasting deficiencies in their physical stature. Other studies have looked at how the effects of malnutrition on the child's health stature may be related to the child's schooling outcomes (e.g., Behrman and Lavy 1994; Glewwe and Jacoby 1995; Glewwe and King 2001; Glewwe, Jacoby and King 2001; Alderman et al. 2001) and largely find that preschool malnutrition has negative effect on a child's school enrollment and academic performance. One of the often stated reasons for this relationship between schooling and early childhood malnutrition (stunting) is that families are unwilling or hesitant to send a physically unfit child to school, in addition to the effect of childhood malnutrition on cognitive development that may be reflected in his/her poor performance or progress at school.

The largely uneducated parents in developing countries, however, may be less likely to recognize the potential correlation between physical fitness and cognitive abilities than they are to recognize the importance of a child's physical strength for family labor. Consequently, parents may end up sending the physically weaker children to school and keep the robust ones for family labor or demand more of their after school time for family labor activities. As a result, studies that ignore the importance of physical stature for child labor (where child labor also matters) may end up with results that understate the effect of malnutrition on enrollment but overstate malnutrition's effect on school performance because it is largely the weaker children with potentially lower cognitive abilities (since malnutrition also hampers child's cognitive development, Dasgupta 1997) who are sent to school. Equity considerations may reinforce the possibility of sending a physically weaker child to school over a stronger sibling if parents feel that the weaker child will have a hard time succeeding in the labor market if he/she doesn't acquire additional skills. Therefore, understanding the role of physical stature of a child in the family's choices between schooling and child labor is not only an important research question in itself but also may help to refine and better understand the observed relationships between childhood malnutrition shocks and academic performance. One issue in using child's physical stature as a covariate in the schooling and child labor equations, however, is that it could be endogenous in both equations because parents might have been making child nutrition decisions in anticipation of specific role for the child. Therefore, an exogenous source of variation in nutritional status that is beyond the control of the parents is needed to identify its effect on schooling and child labor.

In this essay we use two sources of exogenous variation in availability of food (and possibly other amenities) during the critical ages of the child to jointly analyze the effect of early childhood malnutrition on schooling and child labor.[1] First we exploit the natural experiment generated by a massive drought in Ethiopia in 1984 that resulted in a devastating famine that killed about a million people in the country (Jansen, Harris and Penrose 1987). Second, we use the considerable annual fluctuations in rainfall in some localities in the country to identify local weather shocks and the subsequent food deficits in the areas and use these as exogenous sources of malnutrition. In Ethiopia about 85% of the people live on a subsistence agriculture that is almost fully dependent on rainfall conditions. As a result rainfall failures often have big effects on the welfare of households and their members. While grown-ups and older children might also suffer under famines and may sustain some long-term deficiencies in their health and fitness, there is a general consensus in the literature that it is the children at the early years of their life that sustain the biggest long-term damage in their stature and possibly cognitive abilities (Dasgupta 1997). The key purpose of this essay is, therefore, to examine how potential deficiencies in long-term health sustained from early childhood malnutrition are reflected in the child's participation in schooling and family labor.

[1] Porter (2007) analyzed the effect of the 1984 drought shock on the long-term indicators of child nutrition health using data from the first round of the Survey that I'm using. But the first stages of my empirical models in this paper expand her analysis by estimating the effects of localized rainfall shocks on the long-term nutritional status using data from a different cohort of children.

Literature

The determinants of family decisions to send a child to school have been widely investigated in developing countries. Two broad strands of the literature are particularly relevant for the purpose of this essay. The first category includes studies that have examined the role of early childhood malnutrition on child's cognitive abilities, school enrollment and academic performance. Second, there are studies that have jointly analyzed the effect of various child and family characteristics as well as contemporaneous market and environmental shocks on child labor and schooling but did not explicitly address the role played by physical stature (cumulative nutrition outcome) of the child on the family's decision.

The starting point for the studies that examine the effect of early childhood malnutrition on school enrollment and academic performance is whether the former has a lasting effect on the physical and/or cognitive development of a child that could be reflected in his/her schooling outcomes. There appears to be a broad agreement in the nutrition literature (see for example Martorell et al. 1995; Martorell 1997, 1999 for details) that malnutrition during the first three years of a child's life leads to long-lasting deficiencies in the child's physical stature given that height at age three is a strong indicator of height as an adult. While some evidence from the United States indicates that some catch up growth is possible under very favorable conditions (Golden 1994), a recent study by Alderman, Hoddinott and Kinsey (2006) that uses drought and civil war in Zimbabwe as exogenous sources of child malnutrition shows that children who were stunted before their third year will also be shorter as young adults, the period from 12 to 24 months of age being the most vulnerable period for the child. Another recent study by

Maccini and Yang (2008) used rainfall deviations from the mean in specific localities as a shock and found that a rainfall shock during the first year of life is strongly associated with adult height for men and women. This conclusion is based on evidence from those born between 1953 and 1974 in Indonesia, with outcomes measured in the year 2000. The relationship between the timing of a child's malnutrition and his/her cognitive abilities later in life, as measured by scores on intelligence tests, was examined by Glewwe and King (2001) using data from Philippines. They used local price and rainfall variability as instruments for child's nutritional status, as measured by changes in height, and found that malnutrition in the second year of life has the largest impact on scores on a non-verbal intelligence test taken when the child was 8 years old.

The effect of early childhood malnutrition, as measured by the child's age-conditioned height, on school enrollment and educational attainment has been analyzed by several studies by economists. Various approaches have been followed to address the endogeneity of the nutritional status of a child in schooling equations that could arise because of possible simultaneity in family decisions. For example Alderman, Hoddinott and Kinsey (2006) used drought and civil war in Zimbabwe as exogenous sources of child malnutrition in three resettlement villages and found that children who were stunted as preschoolers entered school later and completed less schooling on average. Alderman et al. (2001) used price shocks as identifying instruments for preschool malnutrition and found strong relationship between preschool child nutritional status and subsequent school enrollment in rural Pakistan. Glewwe, Jacoby and King (2001) employed siblings' difference approach and height-for-age of the older sibling as instruments for a child's nutritional status to analyze the relationship between preschool malnutrition and subsequent academic achievement. Using data from Cebu Longitudinal Health and Nutrition Survey from Philippines they find that children who were malnourished as preschoolers enter school later and perform poorly on intelligence tests. Glewwe and Jacoby (1995) used multiple instruments including household wealth proxies, health prices and mother's height and find that early childhood malnutrition resulted in delayed school enrollment in Ghana.

None of these studies, however, addressed the possibility that the estimated relationship between children's physical stature and school enrollment could be understated because of the importance of physical fitness of the child for child labor in the family. In addition, if parents are more likely to send to school a physically less fit child than a more robust child because the latter is better suited for family labor, estimates of academic performance based on those who are enrolled is likely to be biased since the sample will be dominated by malnourished children with potential cognitive deficiencies. The possibility that families may also demand more family labor from the stronger children even after school may further confound the results. Therefore, understanding the importance of physical stature of a child for various child activities including schooling will be one step towards resolving the possible inaccuracies in the estimates for the effect of childhood malnutrition on schooling. One way to do this will be to jointly estimate the effect of the child's physical stature on schooling and child labor.

Joint analysis of family decisions between schooling and child labor has been conducted in the context of family's response to exogenous income and other shocks. For example, Bourguignon, Ferreira and Leite (2003) examined the response of child labor and schooling to conditional cash transfers for poor families with children in Brazil. Using a multinomial logit model for joint analysis of child labor and schooling, they find that the conditional cash transfers substantially increase school enrollment and reduce child labor outside the household for 10 to 15 year olds. Dillon (2008) uses a multivariate probit model to control for cross-equation correlations in analyzing the response of child participation in schooling, home production and market production to household level shocks. He finds that production shocks in the form of pest infestation at the time of harvest lead to child withdrawal from schooling and increased participation in farming in northern Mali. Edmonds (2006), exploits the sharp discontinuity in age eligibility for South Africa's old age pension as quasi-experiment, and finds that anticipated pension income significantly decreases child labor and increases school attendance of 10–17-year-olds. Ravallion and Wodon (2000) also analyzed the joint effect of school subsidies on child labor and schooling and find that the subsidy increases schooling but does not have significant effect on child labor. Similarly, Dammert (2008) and Kruger (2007) exploit coca production shocks in Peru and variations in coffee production in Brazil, respectively, to simultaneously analyze the influence of household income shocks on family decisions towards child labor and schooling. However, all of these studies essentially examined the effect of contemporaneous shocks on child labor and schooling and none of them addressed the consequences of shocks encountered during early childhood on subsequent participation in schooling and child labor. In addition, none of these studies specifically addressed the role of child attributes in influencing the family decisions towards schooling and child labor.

One study that focuses on the role of child attributes is a recent publication by Bacolod and Ranjan (2008) that examines the role of child's cognitive ability in the family's choice between child schooling and child labor using data on child IQ scores in the Cebu Longitudinal Health and Nutrition Survey from Philippines. Applying a multinomial logit model for joint analysis of schooling and child labor, they find that poor households with high-IQ children are more likely to send them to school than poor households with low-IQ children. On the other hand, they find that children with low IQ scores are more likely to be working or staying idle. However, the causality of these relationships is questionable given that they did not use any exogenous shock that could lead to ability differences among the sample children and the IQ tests were not age specific and some of the children already were at school when they took the tests. Although they try to use differences in the IQ scores of siblings in the form of standardized z-scores to control for unobserved household characteristics, it is plausible that unobserved parental preferences could change over time and some of the siblings may have received better preschool inputs from the parents than others. In addition, it is not fully persuasive that ordinary parents would be able to discern much of the ability differences between their children at an early age.

This essay also focuses on the role of one of the child attributes, physical stature, in the parents' decisions on child activities. This is perhaps more appealing attribute to focus on than is ability because differences in physical stature will be easier for the parents to recognize. While it is possible that the child's physical stature as a measure of cumulative nutrition could embody some aspects of the cognitive ability of the child, as demonstrated by Glewwe and King (2001), for example, it is more likely that recognition of the differences in physical stature rather than ability may drive parents' decisions to allocate their children between schooling and various other activities. When parents' decision is influenced by physical stature rather than ability it is possible that parents may actually end up keeping the physically more fit children for child labor and send the physically weaker children to school and if the deficiency in the physical fitness is a result of unfavorable childhood conditions like malnutrition then the physically weaker child will also be less able, leading to bias in estimates when the importance of child labor is ignored. This study will try to address this issue using exogenous sources of variation in the physical stature of children.

Theoretical Framework

The basic research question in this essay can be described in a simple household utility maximization model for a family with one child and unified preferences as in Ravallion and Woodon (2000) and Bacolod and Ranjan (2008) among others. For convenience the child's life is classified into three periods: preschool age, school age and post-school age. In the preschool period, the parents invest in the health of the child in the form of nutrition, health care and other treatments. The health of the child in this period could also be influenced by factors beyond the control of the family like weather shocks and availability of health care services. In the second period parents decide whether to send the child to school or to child labor. In the third period, the child works and earns his/her own income, while parents retire and consume the return on the assets they saved during the earlier periods and possible transfers from their children. The focus here is on the decision problem that parents face in the second period given the outcome of their decisions in the first period.

Assuming that parents are altruistic towards their children and the utility parents derive from own consumption is linearly separable from that they derive from the child's utility as in Barro and Becker (1986), Cigno and Rosati (2005) and Dillon (2008), among others, the parents' utility may be stated as

$$U = \sum_t u_t(c_t^p) + \beta U^*(c_1^c, c_2^c, y_3) \qquad t=1, 2, 3 \tag{1.1}$$

where, c_t^p is parents' consumption in period *t*, U^* is child's maximized utility, c_1^c is child's consumption in period 1 including healthcare, c_2^c is child's consumption in period 2 including healthcare but excluding school expenses, y_3 is child's income in the post-school period and β is a measure of parental altruism towards the child where we assume $0<\beta\leq 1$. Both $u_t(.)$ and $U^*(.)$ are assumed to be quasi-concave and strictly increasing in all of their arguments. In period 2, c_1^p and c_1^c are no longer part of the decision problem of the parents. However, c_1^c determines the child's pre-school stock of human capital in the form of physical stature and cognitive ability, given the child's genetic and natural endowments. And according to the literature on nutrition physical stature at the preschool age (that is also correlated with cognitive ability) is a strong predictor of the later physical stature of the child as previously discussed. Let h_1 denote this preschool physical stature of the child measured in terms of height-for-age. Assuming that the trajectory for the physical human capital of the child is completely set in the preschool age and building on Glewwe (2002), the human capital production function of the child in period 2 may be stated as

$$h = \gamma(h_1, \mu) s(T_c^s, Q) \tag{1.2}$$

where, *γ(.)* is the 'learning efficiency' of the child that depends on the unobserved factors *(μ)* that include genetically inherited ability, child's motivation, etc. as well as the child's physical fitness accumulated during the preschool period (h_1). On the other hand, *s*(.) is the schooling production function that depends on the amount of child's time spent in schooling and studying, T_c^s, and a vector of other educational inputs and school characteristics, Q. In period 2, *γ*(.) is assumed to be predetermined while the interaction between *γ*(.) and *s*(.) produces new human capital. For simplicity we are assuming that accumulation of long-term human capital does not depend on fluctuations in consumption after the preschool period. That is why c_2^c is not included as an argument in human capital production function for period 2.

The human capital the child accumulates through period 2 along with the net parental transfers determines his/her income in the post school period, y_3:

$$y_3 = \omega h - m \tag{1.3}$$

where *m* is the amount of net transfers a child makes to his/her parents in the post school period and *ω* is the return to human capital. Family income in period 2, y_2, comes from three sources. For a typical agricultural household in a developing country like Ethiopia, the principal source of income is family production where both adult and child labor are used as inputs. The other potential sources of income for agricultural households include wage earnings and remittances. Letting w_p and w_c be the opportunity costs of the parent's time and child's time, respectively, the total family income in period 2 is given as,

$$y_2 = q(T_p^f, T_c^f \mid K) - w_p T_p^f - w_c(h_1)T_c^f + w_p T_p^w + R \qquad (1.4)$$

where $q(.)$ is the total value of family production, T_p^f is parent's time in family production, T_p^w is parent's time in wage employment, T_c^f is child's time in family production, K is a vector of family assets like land and livestock, and R stands for family income from other sources including remittances . We are assuming that the child does not engage in wage employment during the school period which is generally true in the rural Ethiopian context. As such, the child's opportunity cost of time in period 2, w_c, is his/her marginal product in family production and it is assumed to depend on the child's physical fitness developed in period 1. In other words, we are assuming that w_c is the return (in period 2) to the physical human capital of the child built in period 1. For simplicity we are also assuming away hired labor and non-family labor although cases of the latter may be observed even in subsistence agriculture mainly because of labor-sharing arrangements. Now, letting p represent a vector of prices for the other educational inputs, the cost function for schooling can be derived following the standard procedure for deriving cost functions (for details see Cigno and Rosati 2005, 31-32). Assuming that the production function for schooling stated under 1.2 is homogenous and twice continuously differentiable, we can minimize the cost of inputs, X, subject to a given level of schooling $\bar{s}$ as

$$\min_{T_c^s, Q} X = w(h_1)T_c^s + pQ \qquad S.T. \qquad s(T_c^s, Q) = \bar{s} \qquad (1.5)$$

This gives us the conditional cost function, $X(s, w_c(h_1), p)$ where the cost of schooling depends on the input prices and the level of schooling. $X(.)$ is assumed to exhibit the standard properties of a cost function. Then, normalizing the price of consumption goods to 1, the budget constraint for period 2 can be stated as,

$$y_2 = c_2^p + c_2^c + X(s, w_c(h_1), p) + A \tag{1.6}$$

where A represents parental savings part of which may be transferred to the child in the post school period and y_2 is given by equation 1.4. In period 3 parents retire and live on the returns from their savings from the earlier period and transfers from the child if m is positive. Therefore, the parent's budget constraint for period 3 can be stated as:

$$c_3^p = rA + m \tag{1.7}$$

where r is return on parental assets. The net parental transfers could be positive if child-to-parent transfers exceed parent-to-child transfers. Substituting 1.7 and 1.3 for c_3^p and y_3 in equation 1.1 respectively, and then substituting 1.2 for h we can rewrite the family's utility function in period 2 as,

$$U = u_2(c_2^p) + u_3(rA + m) + \beta U^*(c_2^c, \omega\gamma(.)s(T_c^s, Q) - m) \tag{1.8}$$

Note that $u_1(.)$ is no longer relevant in period 2 and hence ignored. Assuming that the non-negativity constraints for consumption and parental savings are non-binding and also assuming that the time constraint for both the parents and the child is non-binding so that the Lagrangian multipliers on all these constrains are 0, we can maximize[2] 1.8

---

[2] In writing the maximization problem without the expectations operator, we are assuming that parents face no uncertainty about the values of the third period variables like the return to human capital.

subject to 1.6 to obtain the conditions that determine parental decisions on consumption, savings and time use for themselves and for the child. The Lagrangian function for the maximization problem is,

$$\max_{c_2^p, c_2^c, A, T_c^s, T_c^f, T_p^w, T_p^f, Q} L = u_2(c_2^p) + u_3(rA+m) + \beta U^*(c_2^c, \omega\gamma(.)s(T_c^s, Q) - m)$$
$$+ \lambda[q(T_p^f, T_c^f \mid K) - w_p T_p^f - w_c(h_1)T_c^s + w_2^p T_p^w \qquad (1.9)$$
$$+ R - c_2^p - c_2^c - X(s, w(h_1), p) - A]$$

The first order conditions that are relevant for the purpose at hand are,

$$c_2^p : \qquad \frac{\partial u_2(.)}{\partial c_2^p} - \lambda = 0 \qquad (1.10)$$

$$c_2^p : \qquad \frac{\partial u_3(.)}{\partial c_3^p}\frac{\partial c_3^p}{\partial A} - \lambda = 0 \Rightarrow r\frac{\partial u_3(.)}{\partial c_3^p} = \lambda \qquad (1.11)$$

$$c_2^c : \qquad \beta\frac{\partial U^*(.)}{\partial c_2^c} - \lambda = 0 \qquad (1.12)$$

$$T_c^s : \qquad \beta\gamma(.)\frac{\partial U^*(.)}{\partial y_3}\frac{\partial y_3}{\partial h}\frac{\partial h}{\partial s}\frac{\partial s}{\partial T_c^s} - \lambda\frac{\partial X}{\partial s}\frac{\partial s}{\partial T_c^s} = 0$$

$$\Rightarrow \beta\omega\gamma(.)\frac{\partial U^*(.)}{\partial y_3}\frac{\partial h}{\partial s} = \lambda\frac{\partial X}{\partial h} \qquad (1.13)$$

$$T_c^f : \qquad \lambda[\frac{\partial q}{\partial T_c^f} - w_c(h_1)] = 0 \Rightarrow MP_{T_c}^f(T_c^f \mid K, T_p^f) = w_c(h_1) \qquad (1.14)$$

Condition 1.14 states that the marginal product of the child’s time in family production in period 2 equals the opportunity cost of the child’s time that itself is assumed to depend on the child’s physical fitness accumulated during the preschool period. In 1.13 $\partial X/\partial s$ is the marginal cost of schooling that is henceforth denoted by MC_s and $\partial h/\partial s$ is the marginal productivity of schooling in the production of overall human capital henceforth denoted by MP_s^h. The marginal cost of schooling depends on the level of schooling, the opportunity cost of the child's time and price of other educational inputs. Dividing 1.10 by 1.11 we obtain,

$$MRS^{p}_{c_3,c_2} = \frac{\partial u_2(c_2^p)\big/\partial c_2^p}{\partial u_3(rA+m)\big/\partial c_3^p} = r \tag{1.15}$$

The middle term in 1.16 is the marginal rate of inter-temporal substitution between current consumption and future consumption for the parents ($MRS^{p}_{c_3,c_2}$). The equation states that parents save for their future consumption until the marginal utility of the current consumption relative to their future consumption is equated to return on savings (the interest rate). The analogous condition for the child is obtained by dividing 1.12 by 1.13,

$$MRS^{c}_{y_3,c_2} = \frac{\partial U^*(c_2^c, \omega\gamma(.)s(T_c^s, Q) - m)\big/\partial c_2^c}{\partial U^*(c_2^c, \omega\gamma(.)s(T_c^s, Q) - m)\big/\partial y_3} = \omega\left(\frac{\gamma(.)MP_s^h}{MC_s(s, w_c(h_1), p)}\right) \tag{1.16}$$

The middle term in 1.16 is the marginal rate of inter-temporal substitution between current consumption and future income for the child ($MRS^{c}_{y_3,c_2}$). The term in the parenthesis on the right hand side of this equation may be interpreted as the marginal return to investment in schooling in terms of building the overall human capital of the child. The entire term on the right hand side then represents the marginal return to human capital built through schooling. Note that the effectiveness of investment in schooling in building the overall human capital (knowledge and capability) of the child depends on the learning efficiency of the child and marginal productivity of schooling in the production of human capital. While some of the learning efficiency could be genetic and may be acquired through inheritance, part of it is built through investment in nutrition and healthcare during the preschool period. However, it is assumed that parents treat these as sunk costs when they make decisions about consumption and time use in period 2.

Assuming that parents try to allocate the family's resources so as to maximize the life time utility for themselves and the child and given that total utility is strictly increasing in both the parents' and the child's consumption, they will allocate the child's time between T_c^s and T_c^f by comparing the future marginal return to investment in human capital (given by the right hand side of 1.16) to the return that the child's contribution to the current income could bring in if it were to be saved for future consumption (r). If $r > \omega[\gamma(.)MP_s^h / MC_s]$, then parents are likely to allocate more of the child's time to generating current income through child labor and less to schooling since marginal return to asset savings is greater than the marginal return to human capital. On the other hand, if $r < \omega[\gamma(.)MP_s^h / MC_s]$, then parents are likely to allocate more of the child's time to schooling and less to family work since marginal return to human capital in the future is greater than the marginal return to savings. Therefore, the optimal allocation of the child's time between schooling and current income generating activities is given by,

$$MRS^c_{y_3,c_2} = \omega\left(\frac{\gamma(.)MP_s^h}{MC_s(s, w_c(h_1), p)}\right) = r = MRS^p_{c_3,c_2} \tag{1.17}$$

A situation where a child is full-time student is a discrete case that may arise because of a very high marginal return to investment in schooling relative to the return from savings that could be made from potential contribution of the child to the current income. Similarly, a situation where a child works full time could arise because of a very high return to the child's current contribution to income compared to the anticipated marginal return to schooling. In practice, the possibility of observing these discrete cases is often high due to the fact that schooling requires some minimal level of time commitment from the child and the perfect continuity in time allocation presumed under the solutions above may not hold.

The influence of our key variable of interest, preschool physical fitness (h_1), on the parental decisions about the child's time allocation comes in through its effect on the marginal return to human capital. And h_1 affects the marginal return to human capital through its effect on the marginal cost of schooling, efficiency of learning and marginal productivity of schooling in the production of human capital. For a given ω, therefore, the net effect of a higher value of h_1 on the return to investment in human capital depends on the relative strength of the following two effects.

$$\left(\frac{\partial MC_s}{\partial w_c}\frac{\partial w_c}{\partial h_1}\right) \quad ? \quad MP_s^h\left(\frac{\partial \gamma}{\partial h_1}\right)+\gamma(.)\left(\frac{\partial MP_s^h}{\partial \gamma}\frac{\partial \gamma}{\partial h_1}\right) \qquad (1.18)$$

The term to the left of the question mark in 1.18 represents the effect of h_1 on the marginal cost of schooling. This comes in through the marginal productivity of the child in family production activities. The higher the value of h_1 the more productive the child will be in the family activities and the higher will be the value of his/her w_c. Therefore, a higher h_1 leads to higher marginal opportunity cost of schooling and the sign of the term to the left of the question mark is positive. This tends to reduce the marginal return to investment in human capital. Mathematically, this is easy to see since MC_s is in the denominator of the expression for marginal return to investment in human capital in equation 1.17.

On the other hand, the expression to the right of the question mark in 1.18 represents the effect of h_1 on learning efficiency and marginal productivity of schooling in building human capital. The term $\partial\gamma(.)/\partial h_1$, captures the effect of physical fitness on the learning efficiency of the child that is assumed to be positive because of the empirically observed positive relationship between physical stature and cognitive ability. Note that learning efficiency is important in learning knowledge and skills not only at school but also outside the school environment and $\partial\gamma(.)/\partial h_1$ represents the effect of h_1 on this overall effectiveness in learning knowledge. The second term on the right captures the effect of h_1 on the marginal productivity of schooling in building human capital and this comes in through the effect of h_1 on the learning efficiency. Higher value of h_1 leads to more effectiveness in learning that itself is expected to improve productivity of schooling in building human capital rendering the sign of the entire expression to the right of the question mark to also be positive. Therefore, higher h_1 tends to boost return to investment in human capital through its effect on $\gamma(.)$ and MP_s^h since both of these terms are in the numerator of the expression for the return to investment in human capital stated under equation 1.17.

The net effect of h_1 on the marginal return to investment in human capital will be negative if its effect on MC_s is stronger than its combined effect on $\gamma(.)$ and MP_s^h. For given values of r, ω and parental preferences, therefore, parents will have an incentive to keep a physically stronger child out of school so as to engage in the child labor activities. This means, parents believe that the marginal productivity of such a child in the current family activities is higher than whatever future gains (net of the cost of schooling) in earnings the child could achieve through schooling. On the other hand, if the combined effect of h_1 on the overall efficiency of learning and the marginal productivity of schooling is stronger than its effect on MC_s, parents will have an incentive to send the child to school. Whether parents allow the child to be a full time student by letting him/her to focus on studying even after coming back from attending school or ask him/her to work after school can be established following similar reasoning. This is so because studying after school is part of the human capital building process whose opportunity cost could be measured by the marginal productivity of the child in family activities just like attending school. Therefore, the effect of physical stature of the child on child labor and schooling is theoretically ambiguous as opposed to the prevailing wisdom that it enhances the chances of attending school.

To empirically test the implications of this theoretical model, we need to derive the parental demand functions for own and child's consumption as well as time use. When specific structural forms are assumed for the utility function, specific forms for the demand functions can be derived by simultaneously solving the relevant first order conditions stated above and the budget constraint state under 1.6. For a general form of the utility function assumed here, however, the demand functions will take the following general forms.

$$T_c^{s*} = T_C^s(\omega, h_1, p, R, w_p, m, r, \mu) \tag{1.19}$$

$$T_c^{f*} = T_c^f(\omega, h_1, p, R, w_p, m, r, \mu) \tag{1.20}$$

The demand functions for other choice variables c_2^{c*}, c_2^{p*}, T_p^{f*}, T_p^{w*}, $A*$ and $Q*$ take similar general forms. It is important to note that these demand functions are interdependent because of the simultaneous nature of parental decisions. This is particularly magnified in the case of time use decisions because of the fixed time constraint. For a child constrained with only 24 hours a day, more time for family labor means less time for attending school and studying then after. Therefore, joint estimates of the demand functions will generally provide more accurate estimates of the effects of the covariates on each of the parental choices than the estimates from independent equations for each demand function. This is so because some of the factors that influence parental decisions may not be observable and hence cannot be included as regressors in each equation. As a result the errors that include these unobservables will be correlated across equations and joint estimation techniques that exploit these correlations will lead to more accurate estimates.

To specify such joint empirical models for parental demand for child labor and schooling we first define the indirect utility function for the parents, $v(\omega, h_1, p, R, w_p, m, r, \mu)$, by successively substituting the relevant demand functions into 1.2, 1.3, and 1.7 and the resulting functions into 1.8 along with c_2^{c*} and c_2^{p*}. The indirect utility function is thus defined in terms of observables. From the researcher's perspective, however, there are unobservable elements that may influence parents' decisions and restating the utility function by adding these random components to the indirect utility provides the basis for the empirical models specified in the next section.

Econometric Models and Estimation Methodology

The main purpose of this essay is to analyze the effect of physical stature of a child in the form of height-for-age z-scores on his/her participation in child labor and schooling. The empirical model for the analysis has to allow for the potential correlation between the error terms of the schooling and child labor equations that arises because of the joint nature of the two decisions. In addition, the models have to fit at least two common features of the data on parents' allocation of children's time in developing countries, as already noted by Maitra and Ray (2002), Shafik (2007), and Bacolod and Ranjan (2008) among others. These include a situation where the child neither works nor attends school and a situation where the child combines both. Recognizing these, we can lay down at least four scenarios for parents' possible child activity choices in a given period. First, the child neither works nor attends school; i.e., $T_c^s = 0$ and $T_c^f = 0$. Let's denote parental utility from this choice of child's activity as $v_0(.)$. Second, the child attends school but does not work, $T_c^s > 0$ and $T_c^f = 0$ with utility $v_1(.)$. Third, the child works but does not attend school, $T_c^s = 0$ and $T_c^f > 0$ with utility $v_2(.)$ and the final choice is for the child to combine both school and work, $T_c^s > 0$ and $T_c^f > 0$, with utility $v_3(.)$. If we denote the random components of the parents' utility from the child i's activity j as ε_{ij}, the additive randomized utility (see Cameron and Trivedi 2005, 504) of the parents from child i's activity j can be stated as,

$$U_{ij} = v_{ij}(.) + \varepsilon_{ij}, \quad \text{where, i=1,…,n,} \quad \text{j=0,1,2,3.} \tag{1.21}$$

Parents will choose activity j over activity k for child i if and only if $U_{ij}(.) \geq U_{ik}(.)$ for all j≠k and j,k=0,1,2,3. Denoting the parents' activity alternatives for child i by y_i, the probability that activity j is chosen over k is then, given as,

$$P(y_i = j) = P[U_{ij} \geq U_{ik}], \text{ for all j≠k.}$$

$$= P[v_{ij}(.) + \varepsilon_{ij} \geq v_{ik}(.) + \varepsilon_{ik}], \text{for all j≠k.}$$

$$= P[\varepsilon_{ik} - \varepsilon_{ij} \leq v_{ij} - v_{ik}], \text{ for all j≠k.} \tag{1.22}$$

Assuming that the errors are distributed as type I extreme value and v_{ij} is linear in its arguments, the probabilities in (1.22) can be modeled as multinomial logit with 4 outcomes (see Cameron and Trivedi 2005, 505 and Wooldridge 2002, 497). The multinomial logit model with random effects for child and family level unobserved heterogeneities is described in appendix C. Such a model may be useful to jointly estimate the various child activity equations but does-not fully account for the correlations between errors in the equations because it assumes independence of irrelevant alternatives (IIA). IIA requires that the relative probability of choosing between any two alternatives is unaffected by the presence of a third alternative, but that may not always hold in practice. As an alternative approach, therefore, a bivariate probit model is specified for the child-activity choices described above.

To specify the bivariate probit model for child work and schooling we can define two separate latent variables by adding unobserved random components to the indirect utility parents derive from child schooling and work as,

$$u_{is}^{*} = v_{is}(.) + \varepsilon_{is} \qquad (1.23)$$

$$u_{iw}^{*} = v_{iw}(.) + \varepsilon_{iw} \qquad (1.24)$$

where u_{is}^{*} and u_{iw}^{*} represent additive random utility parents derive from child i's participation in schooling and family work , respectively. Maintaining the assumption that $v_{ij}(.)$ is linear in its arguments, we can restate 1.23 and 1.24 as,

$$u_{is}^{*} = x_{is}^{'}\beta_{s} + \varepsilon_{is} \qquad (1.25)$$

$$u_{iw}^{*} = x_{iw}^{'}\beta_{w} + \varepsilon_{iw} \qquad (1.26)$$

where $x_{ij}^{'}$ represents a vector of covariates including our key variable of interest, physical stature of the child (h_1). The latent variables, u_{is}^{*} and u_{iw}^{*}, are unobserved but let's assume that parents send a child to school or work only when the overall utility from doing so is positive. Then we can define the following dichotomous variables for child's participation in schooling and family work, respectively.

$$s_i = \begin{cases} 1 & if\ u_{is}^{*} > 0 \\ 0 & if\ u_{is}^{*} \leq 0 \end{cases} \qquad (1.27)$$

$$w_i = \begin{cases} 1 & if\ u_{iw}^{*} > 0 \\ 0 & if\ u_{iw}^{*} \leq 0 \end{cases} \qquad (1.28)$$

The four possible choices parents can make regarding child i's time use are now: $s_i=0, w_i=0; s_i=0, w_i=1; s_i=1, w_i=0;$ and $s_i=1, w_i=1$. Assuming that ε_{iw} and ε_{is} are distributed jointly normal with means zero, variances one, and correlation ρ, the probabilities of observing each of these joint outcomes can be specified as bivariate normal. For example, the probability of observing $s_i=1, w_i=1$ can be stated as,

$$\begin{aligned} p_{ik} &= p[s_i = 1,\ w_i = 1] \\ &= p[u_{is}^* > 0,\ \ u_{iw}^* > 0] \\ &= p[\varepsilon_{is} < x_{is}^{'}\beta_s,\ \ \varepsilon_{iw} < x_{iw}^{'}\beta_w] \\ &= \int_{-\infty}^{x_{is}^{'}\beta_s} \int_{-\infty}^{x_{iw}^{''}\beta_w} \phi(z_s, z_w, \rho) dz_s dz_w \\ &= \Phi(x_{is}^{'}\beta_s, x_{iw}^{'}\beta_w, \rho) \end{aligned} \tag{1.29}$$

where $\phi(.)$ and $\Phi(.)$ are the standardized bivariate normal density and the cumulative density function for (z_s, z_w), respectively. We can state similar bivariate cumulative density and density functions for the other three possible outcomes. Following Green (2007), these can be generalized as,

$$\begin{aligned} p_{ik} &= p[s_i = j,\ w_i = k] \\ &= \Phi(\delta_{is} x_{is}^{'}\beta_s, \delta_{iw} x_{iw}^{'}\beta_w, \delta_{is}\delta_{iw}\rho) \end{aligned} \tag{1.30}$$

where the indicator function $\delta_{is}=1$ if $s_i=1$ and $\delta_{is}=-1$ if $s_i=0$. Similarly, $\delta_{iw}=1$ if $w_i=1$ and $\delta_{iw}=-1$ if $w_i=0$. Then the log-likelihood function for the bivariate probit model can be stated as,

$$\ln L = \sum_i \ln \Phi(\delta_{is} x_{is}^{'}\beta_s, \delta_{iw} x_{iw}^{'}\beta_w, \delta_{is}\delta_{iw}\rho) \tag{1.31}$$

We estimate 1.31 and 1.22 using maximum likelihood procedures. We also try to estimate a semi-nonparametric bivariate model for child schooling and labor using the procedure developed in Gallant and Nychka (1987). In their approach, as slightly modified by De Luca (2008), the unknown joint density of the errors is approximated by the Hermite series of the form,

$$h(\varepsilon_s, \varepsilon_w) = \frac{1}{\psi_N} \alpha_r(\varepsilon_s, \varepsilon_w)^2 \phi(\varepsilon_s)\phi(\varepsilon_w) \qquad (1.32)$$

where, *ϕ(.)* is the standardized Gaussian density, $\alpha_r(\varepsilon_s, \varepsilon_w) = \sum_{i=0}^{r_1} \sum_{j=0}^{r_2} \alpha_{ij} \varepsilon_s^i \varepsilon_w^j$ is a polynomial in ε_s and ε_w of order $r=(r_1,r_2)$ and, $\psi_N = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \alpha_r(\varepsilon_s, \varepsilon_w)^2 \phi(\varepsilon_s)\phi(\varepsilon_w) d\varepsilon_s d\varepsilon_w$ is a normalization factor that ensures h(.) integrates to 1. Equation 1.32 approximates the joint density of the errors as the product of a squared polynomial and a standardized bivariate normal density where the latter is assumed just for convenience. Gallant and Nychka (1987) demonstrate that 1.32 approximates densities with arbitrary skewness and kurtosis except those that are violently oscillatory. In implementation, the vector of parameters $\alpha = (\alpha_{00}, \alpha_{01}, ..., \alpha_{r_1 r_2})$ is normalized by setting $\alpha_{00} = 1$ since the polynomial expansion in 1.32 is invariant to multiplication of the parameter vector by a scalar. The specification of the pseudo–log-likelihood function and the detailed procedures for implementation of the model are explained in De Luca (2008). This approach not only relaxes the parametric assumption of the bivariate probit model in estimating the coefficients but also allows detailed examination of the characteristics of the error densities for different values of r_1 and r_2.

In addition to the child's height-for-age z scores as a measure of the child's physical fitness, the vector of covariates in all the models includes child's age and gender, number of siblings, livestock and land area owned as measures of the household's wealth status, parents' age and education, as well as distance to a primary school as a proxy for cost of schooling. The indicators for household wealth could be thought of as proxies for household income, discussed in the theoretical model. Information on household income gathered through surveys in the rural areas of developing countries is often unreliable and wealth indicators could be better measures of household well-being. Controlling for wealth indicators is important because the need for child labor and the ability of the families to send their children to school could vary with wealth status. Variation across households and changes over time in wealth indicators could also be correlated with nutritional status of children; thus failing to control for wealth indicators could bias our estimates. The theoretical model described above also implies that the wage rate for child labor is a relevant variable that should be accounted for in the empirical model since the wage paid to a child could be correlated with physical stature. However, child labor in rural Ethiopia almost entirely consists of unpaid family labor, so information on formal wage rates for children is unavailable. The child's opportunity cost of time is essentially his/her marginal product in the family production activity and to the extent that the marginal productivity depends on having other assets to work with, children in the families with more land and livestock could have higher opportunity cost of time than children with less assets. Therefore, inclusion of land and asset ownership as covariates may partly control for the opportunity cost of the child's time.

The vectors of coefficients from the bivariate probit and multinomial logit models are used to calculate the marginal effects of the covariates on the probability of observing each of the joint outcomes: $p(s_i=0, w_i=0)$, $p(s_i=0, w_i=1)$, $p(s_i=1, w_i=0)$, and $p(s_i=1, w_i=1)$. For the purpose of comparison with other studies that estimated an independent equation just for schooling, we also estimate the standard probit models for the child's school attendance and participation in family work. Therefore, the marginal effects of the covariates on $p(s_i=1)$ and $p(w_i=1)$ are computed using both the joint models as well as independent probit models. As briefly described in the previous section, the marginal effect of our key variable of interest, child's physical stature on child schooling and child labor is theoretically ambiguous. The existing literature generally argues that better physical fitness enhances the chances that a child attends schooling implying that its effect on $p(s_i=1)$ will be strongly positive. The effects of physical fitness on the joint outcomes have not been examined by the existing studies. Therefore, the estimates here help us to answer an important question of whether child's physical fitness enhances the child's chances of being a full time student, $p(s_i=1, w_i=0)$, or part-time student, $p(s_i=1, w_i=1)$, or even full-time worker, $p(s_i=0, w_i=1)$.

One important issue that needs to be addressed in estimating these models is the potential endogeneity of the child's physical stature in both schooling and child labor equations. Endogeneity could arise because parents may be providing preferential treatment in terms of nutrition to some children (particularly when resources are limited) in anticipation of specific role for each child depending on their perceptions regarding the importance of physical fitness for each of the child's anticipated roles. For example, parents may feed the oldest child very well so that he/she quickly grows up and helps them in fulfilling the family labor needs. If this is the case it may be the anticipated role for the child (schooling or labor) that is determining his physical stature rather than the other way round and the estimates may not represent a causal effect. Therefore, an exogenous source of variation in nutrition status that is beyond the control of the parents is needed to identify its effects on schooling and child labor. Exposure to a famine caused by a massive drought and localized rainfall shocks are used as identifying instruments as discussed in the next section.

Another critical issue is how to implement instrumental variables estimation in the context of these heavily nonlinear models for non-binary outcomes. There are at least three approaches that have been used to address this issue in various contexts. One possibility is to jointly estimate the first stage equation for the endogenous variable and the second-stage equation for the outcome variable of interest, for example, using the full information maximum likelihood approach to obtain asymptotically efficient estimators as initially proposed by Hausman (1975) . However, the application of this method generally depends on some arbitrary assumptions about the joint distribution of the errors in the two equations the validity of which cannot be readily verified.

The other commonly applied method is what may be called ‘two-stage predictor substitution’ (2SPS) where the endogenous regressor in the second-stage equation is replaced by its predicted value from a separately run auxiliary regression correcting the standard errors for the resulting measurement error bias (for some of the recent applications of this method see Lu and McGuire 2002; Meer and Rosen 2004; Savage and Wright 2003; Gramm 2003). Unlike the linear models where the two-stage predictor substitution leads to consistent estimates, however, the consistency of such estimates in the non-linear context has not been well established. In fact Terza, Basu and Rathouz (2008) show that such a method generally leads to inconsistent estimates in the non-linear models. On the other hand, they demonstrate that an alternative method that requires inclusion of the residual from the first-stage auxiliary regression in the second-stage equation provides consistent estimates. The two-stage residual inclusion (2SRI) method has been recently used by a number of empirical studies (see Stuart, Doshi, and Terza 2009; Shea et al. 2007; Gibson et al. 2006; Shin and Moon 2007; DeSimone 2002; Baser et al. 2004) but its theoretical properties in such applications have not been formally examined until the latest work by Terza, Basu and Rathouz (2008).

According to Terza, Basu and Rathouz (2008) the 2SRI method provides consistent estimates because the unobserved factors that led to endogeneity of the regressor can be controlled for by the residuals from the first stage auxiliary regression as long as we can find valid identifying instruments. This method provides not only consistent estimates but asymptotically correct standard errors. They test their theoretical results about the consistency of the 2SRI and inconsistency of 2SPS estimates using simulated data with 5,000 and 20,000 observations. They find negligible biases in the 2SRI estimates and several times larger biases in the 2SPS estimates for a duration model with multinomial endogenous treatments and ordered logit model with count-valued endogenous treatments. They apply the two methods to actual data as well and find that the 2SPS method substantially overestimates the effect of the endogenous variable. Therefore, we use the 2SRI method to address the potential endogeneity of the child's physical stature in the bivariate probit and multinomial logit models for child labor and schooling where the first stage is a linear model for the child's height-for-age z scores.

The two-stage approach fits the models here conceptually as well because parental decisions are formulated as sequential where the early period focuses on building the physical fitness of the child through nutrition and health services and the subsequent period largely focuses on allocating the child's time to schooling or family labor or both.

Identification Strategy

The findings in the literature on nutrition indicate that there is strong relationship between height-for-age in early childhood and height-for-age later in life (e.g., See Martorell et al. 1995; Martorell 1999, 1997). In fact Martorell et al. conclude that "regardless of the choice of reference population, growth is markedly retarded only in early childhood; adolescence is not a period when growth is significantly constrained" (p.1060S). This implies that factors that significantly affect the child's nutritional status during early childhood are likely to be strongly correlated with the child's cumulative nutrition outcome, say height-for-age, later in life. Therefore, if one could find exogenous shocks that could substantially influence the child's nutrition during early childhood, these shocks must be correlated with the child's cumulative nutrition outcomes later in life and hence can be used to identify the effect of the latter on other outcomes for the child like schooling and child labor. Using contemporaneous shocks in such contexts may not be appropriate because they may influence the schooling and child labor outcomes directly, for example by putting the household under resource pressure. On the other hand shocks that happened well in the past are less likely to be directly correlated with current child labor and schooling outcomes except through their long-lasting effect on the child's physical and cognitive abilities.

The fact that the livelihoods of the rural communities in Ethiopia are highly dependent on rainfall conditions provides an opportunity to use rainfall related shocks to identify the effects of early childhood malnutrition on child outcomes later in life. Two approaches are followed in using the rainfall related shocks for this purpose. First, an attempt is made to exploit a famine caused by a massive drought in 1984 where the average rainfall nationwide was 22% below the long-term average, making it the worst drought since rainfall data started to be systematically recorded in 1961 (Webb, von Joachim and Yohannes 1992). While household level data on experience during the famine are largely unavailable, in 1995 a sample of 1477 households from 15 different sites in the country were asked to recall the three biggest droughts over the previous 20 years in which they lost a substantial amount of their harvest and/or livestock. Nearly half the households reported to have lost substantial crop harvest and/or livestock because of the drought in 1984/85 agricultural seasons. The ages of the children in these sample households could be traced back to the time of the drought to identify the group of children who were particularly vulnerable (1 to 3 years old according to the literature on nutrition). These potentially affected children would have been 10 to 12 years old in 1994.

The interactions between dummy variables that identify these children and a dummy variable that identifies households who reported to have faced a substantial shock at the time are used as the first set of identifying instruments for early childhood malnutrition. That is, the identifying instruments are generated by interacting a dummy for the reported household level shock with a dummy for being age 1, a dummy for being age 2 and a dummy for being age 3 in 1984. Children who were 4 to 6 years old at the time of the drought (13 to 15 years old in 1994) are included as controls. These are children who must have been less vulnerable at the time of the drought and must have not sustained substantial damage in their physical stature from the shock.[3] Because of the observed linearity in the relationship between height-for-age in early childhood and later in life (Martorell et al. 1995), the age-shock interactions correlated with height-for-age in the early childhood period should be correlated with height-for-age in 1994 and the subsequent periods. To control for the genetic variation in height we also include the mother's and father's height as additional covariates in the first stage regressions for child's height-for age.[4] This approach is implemented using data from the first round of the Ethiopian Rural Household Survey (ERHS) conducted in 1994 and another round in 1995.

In the absence of detailed data on household experience at the time of the drought, however, the famine shock may still be an imperfect way to accurately identify the degree of malnutrition faced by children from different households. This is so because the capabilities of the households to cope with crop and livestock loss might differ. Another issue with using the famine shocks to identify the effects of malnutrition is that children who survived the famine and are found alive in 1994 could be the stronger ones who could withstand the effects of the drought, while weaker children might have already

[3] Children who were born at and after 1984 may not be an effective comparison group because they may also have been the victims of the after-effects of the drought at their critical age. These children, therefore, are excluded from the sample.
[4] Mother's height was used for similar purpose by Glewwe and Jacoby (1995).

died, in which case the effect of the shock could be understated.[5] Another concern with this approach is that parents' age recalls may entail some errors in a situation where formal records of child's birth date are not kept, as is largely true in rural Ethiopia. This may be a more serious problem particularly when age recalling involves longer time periods.

As a way of validating our results from the 1984-drought based identification strategy, therefore, an alternative strategy based on localized rain-fall shocks is implemented using data from a different cohort of children who were 1 to 6 years old at the time of the first round of the survey in 1994. The fact that the birth dates for these children are relatively close to the survey period is expected to make it easier for the parents to accurately recall the child's age and hence minimize the potential age-recall error bias. The localized rainfall shocks are defined on the basis of the deviations of the annual rainfall in the locality from its long-term mean.[6] Both substantial rain deficits and excessive rains are considered rainfall shocks since both can lead to crop failure. Substantial rain deficit is represented by a dummy that takes a value of 1 if the rainfall shortfall from the long-term mean is bigger than 1 standard deviation and excessive rainfall shock is represented by a dummy taking a value of 1 if the excess of rain over the long-term mean exceeds 1 standard deviation. Because of the erratic nature of rainfall in most localities in Ethiopia, the long-term standard deviations of rainfall are quite large representing more than 15% of the mean annual rainfall on average. Therefore, rainfall deficits and excesses exceeding 1 standard deviation represent substantial shock that may

[5] But the data on mortality history gathered during the 1995 round of the survey don't show any unusually high mortality in 1984 for the age group included in our sample.
[6] A similar strategy was followed by Maccini and Yang (2009).

lead to crop failures and significant reductions in consumption in rural Ethiopia. For example, Dercon (2002) finds that a 10% decrease in rainfall from the long-term mean decreases food consumption by up to 5% and localized rainfall shortfalls of this magnitude or bigger are quite common in Ethiopia.

Therefore, the rainfall shocks faced by a child during the first 3 years of life are taken as exogenous indicators of early childhood malnutrition and hence used as instruments for the child's age-standardized heights in the child labor and schooling models. In this case height-for-age measured towards the end of the preschool period are used since the anthropometric data were gathered for all members of the sample households in 1994, 1995 and 1997. The genetic variations in children's height are controlled for by mother's and father's heights in this approach as well. Malnutrition induced by exogenous rainfall shocks is expected to explain what is left of these natural differences in the heights of children. The schooling and child labor models for this cohort of children are estimated using data from the latest two rounds of the survey conducted in 1999 and 2004. The age range for this cohort in 2004 is similar to the age range for the older cohort in 1994. Therefore, results from the two identification strategies are expected to be at least qualitatively comparable although rainfall shortfalls might be weaker instruments than the major famine shock.

Data and Summary Statistics

The analysis in this essay is based on data from the various rounds of the Ethiopian rural household survey (ERHS) conducted by the Economics Department of

Addis Ababa University in collaboration with the Center for the Study of African Economies at the University of Oxford, the International Food Policy Research Institute (IFPRI) and the US Agency for International Development (USAID). ERHS is a unique longitudinal data set in Ethiopia the first round of which was conducted in 1994 (subsequently referred to as 1994a) and covered 1477 households from 15 different sites across the country. Another round was conducted later in 1994 (henceforth referred to as 1994b) followed by one round each in 1995, 1997, 1999 and 2004. The attrition rate was small between successive rounds and the 6th round in 2004 managed to successfully re-interview about 1370 of the households in the original sample. The 15 sites (called peasant associations) were selected to represent the major farming systems[7] in the county and households were randomly selected from the list of households in each peasant association. While strictly speaking ERHS is not nationally representative[8], the major statistics from this survey are very close to those from nationally representative surveys (see Dercon 2000).

All the rounds of the ERHS data contain detailed information on household demographics, asset ownership, as well as income and consumption. Information on height and weight for all household members was gathered in all the rounds except in 1999. The anthropometric data in the ERHS are directly collected by the enumerators using measuring scales. While this may not totally eliminate measurement errors, it is expected to minimize it compared to the surveys where data on respondent heights and weights are collected through self-reporting. Information on exposure to significant

[7] These are the grain-plough areas of the Northern and Central highlands, the Enset-growing areas and the sorghum-hoe areas.
[8] The pastoralist farming system was not represented,

drought shocks was gathered during the 1995 round. In this round households were asked to list three most important droughts (listed in the order of severity) over the last 20 years because of which they suffered substantial loss of harvest and/or livestock.

The analysis that uses the 1984 drought as exogenous source of malnutrition focuses on the cohort of children who were 10 to 15 years old during the 1994a round (henceforth called the older cohort) who must have been 1 to 6 years old during the 1984 drought. Those who were age 1 to 3 may be considered as the treatment group because this is the age range that evidence from the nutrition literature shows is the critical period where malnutrition can have a lasting impact on the child's stature. Those who were 4 to 6 could be considered as the comparison group because there is not strong evidence that malnutrition beyond age 3 has a lasting impact on the child's physical stature. For the analysis where localized rainfall shocks are used as exogenous sources of malnutrition data from the cohort of children who were 1 to 6 years old during 1994a round (henceforth called the younger cohort) are used.

Data on child activities were collected in 1994a, 1995, 1999 and 2004. Child activity data for the analysis involving the older cohort comes from 1994a and 1995 rounds. However, the level of detail in the data on child-activity was different in the two rounds. In 1994a, data on child activities were collected as part of main activities for all household members and the main activity categories for children included student, farm worker, domestic worker, domestic and farm worker, off-farm business worker, and not involved in work[9]. This round did not ask questions on activity combinations of children.

---

[9] While some of the activities such as farming could vary seasonally, most of the activities in which children participate like herding cattle, fetching water and fuel wood, watching the little kids and other

On the other hand the 1995 round collected data on not only the main activity of the child but also on secondary and tertiary activities. Specifically, the 1995 round asked the 1st, 2nd and 3rd activity of the child ranked in terms of hours spent on each. These activity combinations were collected for both students and non-students. As a result, it is possible to identify children who combined schooling and child labor in 1995 but not in 1994a. Child activity data for the analysis involving the younger cohort comes from the 1999 and the 2004 rounds. Both rounds collected data on both main and secondary activities of all household members including children out of which data on activity combinations for children in the sample cohort are compiled.

Height-for-age z-scores for children were calculated using the software, ANTHRO[10], which uses in-built median heights and weights for similar age groups and gender from the healthy U.S. population as references. The age-standardized height for each child thus represents the number of standard deviations by which the child's height deviates from the median height of the healthy U.S. children with similar age and gender. For the older cohort age-for-height z-scores from 1994a and 1995 rounds are used. An ideal data for the purpose at hand would have been to use height-for-age data collected after the critical period (age 3) but before the school age[11] since the height of the child in this period will fully reflect the outcome of his/her early childhood nutrition experience. Unfortunately, such data are unavailable for the older cohort but the analysis based on

domestic chores are year round activities and there will always be something for children to do throughout the year. Therefore, seasonality is assumed away in our analysis.

[10] The software is provided by WHO and is available at http://www.who.int/childgrowth/software/en/index.html, accessed April, 2009.

[11] While there is no official school starting age in Ethiopia, it is rare for a child in rural Ethiopia to start school before age 7 because of the long distances children have to travel to get to the nearest elementary school.

child heights measured in 1994 and 1995 but identified through a malnutrition shock experienced during the early childhood period will still be informative because of the observed linear relationship between height-for-age at the end of the critical period and height-for-age later in life.

On the other hand, data on the preschool height and weight are available for the younger cohort. Therefore, the analysis involving data from the younger cohort uses child height-for-age measured after the critical period but before the school age. For those who were 4 to 6 years old during 1994a, height data reported in 1994a or 1994b (if height is missing in 1994a) are taken. For those who were 3 years old during 1994a, height data reported in 1995 round are taken while for those who were 1 or 2 years old during 1994a, height data reported in 1997 are taken. Therefore, estimation results from the younger cohort are expected to directly reflect the effects of early childhood malnutrition on the child activity choices.

The monthly data on rainfall for the stations closest to the survey sites were obtained from the Ethiopian Meteorological Agency for the period from 1970 to 2006. The key rainfall data needed for the purpose at hand were for the 8 years or 96 months from 1988-1995 for each of the 15 sites when the children in the younger cohort were at their critical stage of development[12]. From the total of these1440 key monthly rainfall records, however, 249 were missing [13](see tables B1&B2 in appendix B for details) and

---

[12] For those who were 1 year old during 1994a round the critical years were taken to be 1993, 1994 and 1995. For those who were 2 years old the critical years were 1992, 1993 and 1994. For the 3 year olds the critical years were 1991, 1992 and 1993. For the 4 year olds the critical years were 1990, 1991 and 1992. For the 5 year olds the critical years are 1989, 1990 and 1991. For the 6 year olds the critical years are 1988, 1989 and 1990.

[13] While this is a lot of missing data by any standard and could possibly lead to understatement of the effects of the rainfall shocks, our results remain nearly unchanged when we re-estimate our models for

replaced by the long-term average for the same month from the same station. The annual rainfall data were then obtained by adding up the monthly data for each year. Annual rainfall deviations for each locality were calculated by subtracting the long-term mean rainfall for the locality from the annual rainfall. Then, three variables representing rainfall deviation that prevailed during the 1st, 2nd and 3rd years of each child in the younger cohort were defined. Three dummies identifying substantial rain-deficit during the 1st, 2nd and 3rd years of the child are then defined to take a value of 1 if the absolute value of the rain deviation for the respective year was greater than 1 long-term standard deviation for the rainfall in the locality. Three other dummies identifying excessive rain during 1st, 2nd, and 3rd years are also defined to take a value of 1 if the rainfall deviation the child faced during the respective year was greater than 1 standard deviation. These six dummies represent the local rainfall shocks[14] that children in the younger cohort experienced during the critical period of their development.

In addition to the child's height-for-age z-scores, a number of control variables are included in the estimated econometric models reported in the next section. These include land and livestock ownership as well as the distance to the nearest primary school. Data on agricultural land area owned by the household were collected in local units that varied across survey sites. The land areas measured in local units were converted into hectares using the land conversion units gathered through the community

---

the younger cohort by excluding all the major cases with missing rainfall data as we report in the next section. Glewwe and King (2001) also used rainfall data with large number of missing observations as an instrument for child malnutrition in Philippines and pointed out that the instrument could have understated the effects of child malnutrition on cognitive development.

[14] The identification strategy based on the localized rainfall shocks assumes that the households lived at their current site for at least the first 6 years of the child's life. According to the data collected on the migration history of the household head and his/her spouse during the 1994b round, the household head was either born in the survey site or arrived before 13 years except 2 cases where the head arrived before 7 years and 5 years. Therefore, mobility doesn't seem to be an issue in our sample.

questionnaire of the ERHS. The various types of livestock owned were also converted into equivalent units and aggregated using the tropical livestock equivalent units that are available in the 1999 round of the survey. Data on distance to the nearest primary school were gathered only in the 1997 and 2004 rounds. Therefore, the distances to primary schools for the 1994a and 1995 rounds are approximated by the distances observed in 1997. The distances to primary schools in 1999 were also approximated by the distances observed in 1997 except when the data gathered in 2004 indicated that a closer school was constructed between 1997 and 1999 in which case the distance information for 1999 were updated to the latest.

The summary statistics for child activities and the covariates used in the first and second stages of the econometric models for the older cohort are presented in table 1 below. In the sample of households interviewed for the 1994a round, there are 1232 children of the older cohort with complete information for the variables of interest. About 24% were students whereas 69% were participating in family labor activities full-time. About 7% were neither working nor attending school. For this round we do not have information as to who among the students were combining work with schooling. On the other hand, 1116 children of the older cohort have information for the variables of interest in the data for the 1995 round out of whom 25% were full time students and 9% were combining schooling and family work. The proportion of students is 10 percentage points higher during the 1995 round. The rapid change may have to do with the aggressive primary school expansion program initiated by the new government at the time. We observe similarly rapid growth in the percentage of students between 1999 and 2004 for the younger cohort.

Table 1. Summary Statistics for the Variables Used in the Econometric Models for the Older Cohort

		1994a (Obs=1232)		1995 (Obs=1116)	
Variable	Description	Mean	St.dev	Mean	St.dev
Student	dummy=1 if student at school	0.24	0.43	0.34	0.48
Working	dummy=1 if working			0.89	0.31
Neither a student nor working	dummy=1 if neither student at school nor working	0.07	0.26	0.02	0.13
Student Only	dummy=1 if student at school and not working			0.09	0.29
Work only	dummy=1 if working and not student at school	0.69	0.46	0.64	0.48
Student and working	dummy=1 if student at school and working			0.25	0.44
Child activity	activity=0 if idle, =1 if student only , =2 if work only, =3 if student and working			2.13	0.63
Main activity of child	main activity=0 if idle, =1 if student, =2 if working	1.61	0.62	1.62	0.51
Sex	sex=1 if male	0.51	0.50	0.51	0.50
Age	Age	12.40	1.76	13.10	1.63
Agri. land area owned	Agricultural land area owned	1.87	1.76	2.12	2.15
Livestock units owned	Tropical livestock units owned	4.89	6.68	4.77	6.35
Father's education	dummy=1 if child's father has completed at least primary education	0.08	0.27	0.08	0.28
Mother's education	dummy=1 if child's mother has completed at least primary education	0.02	0.14	0.02	0.13
Distance to primary school	Distance to the nearest primary school in kilometers	6.00	4.27	6.16	4.40
Household size	Household Size	8.26	3.24	8.73	3.39
Number of siblings	Number of siblings of the child	5.04	2.85	5.28	2.93
Age of father	Age of child's father	49.30	9.82	49.93	9.93
Age of mother	Age of child's mother	40.05	8.43	40.81	8.61
Sex of household head	dummy=1 if h.hold head is male	0.82	0.39	0.81	0.39
Drought Affected in 1984	dummy=1 if household lost crop and/or livestock because of 1984 drought	0.60	0.49	0.62	0.48
First Year in 1984	dummy=1 if child was in his/her first year in 1984	0.20	0.40	0.23	0.42
Second Year in 1984	dummy=1 if child was in his/her				

	second year in 1984	0.14	0.35	0.17	0.37
Third Year in 1984	dummy=1 if child was in his/her third year in 1984	0.19	0.39	0.21	0.41
Drought Affected at 1st Year	dummy=1 if child was in 1st year IN 1984 and belonged to drought affected household	0.12	0.32	0.14	0.35
Drought Affected at 2nd Year	dummy=1 if child was in 2nd year in 1984 and belonged to drought affected household	0.09	0.28	0.11	0.31
Drought Affected at 3rd Year	dummy=1 if child was in 3nd year in 1984 and belonged to drought affected household	0.12	0.33	0.14	0.34
Height-for-age	Child's Height-for-age z-scores	-1.96	1.54	-2.12	1.44
Height-for-age(Dr84)	Obs.94=742, Obs.95 =697	-1.99	1.52	-2.11	1.47
Height-for-age (No Dr84)	Obs.94=490, Obs.95 =419	-1.91	1.55	-2.15	1.38
Height-for-age(Dr84, A1-3)	Obs.94=406, Obs.95 =426	-1.93	1.60	-2.05	1.48
Height-for-age (No Dr84, A1-3)	Obs.94=259, Obs.95 =249	-1.75	1.61	-1.98	1.39
Height-for-age(Dr84, A4-6)	Obs.94=336, Obs.95 =271	-2.07	1.44	-2.21	1.65
Height-for-age (No Dr84, A4-6)	Obs.94=231, Obs.95 =170	-2.09	1.42	-2.39	1.35
Height of mother	Height of mother in centimeters	156.7	7.31	156.8	5.71
Height of father	Height of father in centimeters	168.2	5.35	165.9	7.84

Source: Ethiopian Rural Household Survey.
Notes: Dr84 identifies children affected by the 1984 drought; A1-3 identifies children who were 1 to 3 years old at the time of the 1984 drought while A4-6 identifies children who were age 4 to 6 at the time; Obs.94/Obs.95 stand for the number of observations in the category during the 1994a/1995 rounds.

The average height-for-age z-score for the older cohort is -1.96 during the 1994a round and -2.12 during the 1995 round. This means that children in this cohort are about 2 standard deviations shorter on average than the healthy American children of the same age. According to the WHO standards[15], children with height-for-age z-score less than -2.00 are considered stunted (display retarded growth). About half (49% in 1994a and 53% in 1995) of the children in this cohort were stunted. The evidence in table 1 also

[15] See the WHO growth standards at http://www.who.int/childgrowth/standards/en/, accessed April, 2010.

shows that about 60% of the children in this cohort belonged to households that lost substantial amount of crops and/or livestock because of the 1984 drought out of which well over one half were at the critical age (age 1 to 3) at the time of the drought. There are also some indications that those who were affected by the drought at their critical age were more stunted than children of the same age who were not affected by the drought. According to the height measurements from the 1994a round for example, children affected by the drought at their critical age had average height-for-age of -1.93 compared to -1.75 for children of the same age who were not affected by the drought. The pattern is similar in 1995 as well although the difference is smaller in the latter case and the standard errors are a bit large in both cases perhaps because of small sample sizes for each category. The first stages of the econometric models reported in the next section formally estimate the effect of the drought on height-for-age z-scores.

The summary statistics for the variables used in the econometric models for the younger cohort are presented in table 2 below. Out of the 1184 children in this cohort with complete data for all the variables of interest during the 1999 round, 14% were full-time students while 18% were combining schooling and family work for a total of 32% participation in schooling. About 21% were neither working nor attending school while 48% were full-time participants in family activities. In 2004 there were 1057 children of this cohort with complete information of which 70% were students (13% attending fulltime and 57% combining schooling and work). Again we observe rapid increase in school participation between 1999 and 2004.

Table 2. Summary Statistics for the Variables Used in the Econometric Models for the Younger Cohort

		1999 (Obs=1184)		2004 (Obs=1057)	
Variable	Description	Mean	St.dev	Mean	St.dev.
Student	dummy=1 if student at school	0.32	0.47	0.70	0.46
Student Only	dummy=1 if student at school and not working	0.14	0.35	0.13	0.33
Working	dummy=1 if working	0.65	0.48	0.86	0.34
Work only	dummy=1 if working and not student at school	0.48	0.50	0.29	0.45
Student and working	dummy=1 if student at school and working	0.17	0.38	0.57	0.49
Neither student nor working	dummy=1 if neither student at school nor working	0.21	0.40	0.01	0.12
Main activity of child	main activity=0 if idle, =1 if student, =2 if working	1.27	0.78	1.28	0.48
Child activity	activity=0 if idle, =1 if student only, =3 if work only,=3 if student and working	1.62	1.00	2.42	0.76
Child's Height-for-age	Child's Height-for-age z-scores	-2.22	2.06	-2.25	2.04
Sex	sex=1 if male	0.50	0.50	0.52	0.50
Age	Age	8.24	1.77	13.24	1.77
Household size	Household Size	8.83	3.49	7.20	2.35
Number of siblings	Number of siblings of the child	4.45	2.08	4.66	2.14
Sex of household head	dummy=1 if h.hold head is male	0.81	0.39	0.79	0.41
Father's education	dummy=1 if child's father has completed at least primary education	0.17	0.38	0.18	0.38
Mother's education	dummy=1 if child's mother has completed at least primary education	0.05	0.22	0.06	0.23
Distance to primary school	Distance to the nearest primary school in kilometers	4.69	3.67	3.74	3.04
Age of father	Age of child's father	47.96	10.97	52.81	11.01
Age of mother	Age of child's mother	37.92	8.42	42.80	8.32
Agri. Land area owned	Agricultural land area owned	1.46	1.40	1.44	1.70
Livestock units owned	Tropical livestock units owned	4.08	4.03	6.21	7.57
Rainfall deviation at 1st year	Deviation of rain from long run local mean during 1st year	21.89	171.13	22.50	172.69
Rainfall deviation at 2nd year	Deviation of rain from long run				

	local mean during 2nd year	-3.66	160.23	-2.51	152.11
Rainfall deviation at 3rd year	Deviation of rain from long run local mean during 3rd year	-23.58	169.76	-21.53	164.68
Substantial rain def. at 1st year	dummy=1 if rain deficit at 1st year exceeds local st.dev	0.13	0.33	0.12	0.32
Substantial rain def. at 2nd year	dummy=1 if rain deficit at 2nd year exceeds local st.dev	0.17	0.38	0.16	0.37
Substantial rain def. at 3rd year	dummy=1 if rain deficit at 3rd year exceeds local st.dev	0.20	0.40	0.19	0.39
Substantial rain sur. at 1st year	dummy=1 if rain surplus at 1st year exceeds local st.dev	0.13	0.33	0.13	0.33
Substantial rain sur. at 2nd year	dummy=1 if rain surplus at 2nd year exceeds local st.dev	0.12	0.32	0.11	0.32
Substantial rain sur. at 3rd year	dummy=1 if rain surplus at 3rd year exceeds local st.dev	0.09	0.28	0.09	0.29
Height of mother	Height of mother in centimeters	156.57	6.44	156.45	6.56
Height of father	Height of father in centimeters	166.23	7.90	166.20	7.80

Source: Ethiopian Meteorological agency for the rainfall data, Ethiopian Rural Household Survey for all the Other Variables.

The average pre-school height-for-age z-score for the younger cohort was about -2.2 indicating that stunting of children in Ethiopia is not limited to children who suffered under unusual environmental shocks but rather a widespread phenomenon that afflicts children of all ages. In fact, about 59% of the children in this cohort were stunted during the preschool period and one of the principal causes of stunting is early childhood malnutrition. And malnutrition in most localities in Ethiopia is caused by rainfall fluctuations and the resulting crop failure and livestock death. The large average standard deviation reported for annual rainfall in table 2 is indicative of the degree of unpredictability of rainfall in some of the regions covered by the ERHS survey. Because of this unpredictability at least some of the children born in any given year are likely to face some major crop failure in their locality during their critical years.

The evidence in table 2 shows that a sizable proportion of the children in the younger cohort faced substantial rainfall deficits and/or excessive rains during their first, second or third years. About 12%, 16% and 19% experienced substantial rain deficits in their 1st, 2nd and 3rd years, respectively. On the other hand, 13%, 11% and 9% experienced excessive rains during their 1st, 2nd and 3rd years, respectively. Both substantial rain shortages and excessive rains are considered a shock because farmers develop their cropping patterns on the basis of their expectations about rainfall in their locality that is often based on their individual and collective experience over so many years. Therefore, any variation in rainfall that falls within its long-term standard deviation will generally be anticipated by the farmers but rainfall deficits and surpluses exceeding the long-term standard deviation will be unanticipated and are likely to lead to crop failures. The effects of these early childhood rainfall shocks on the cumulative nutritional health of the children are estimated in the first stage of the econometric models for the younger cohort.

Estimation Results

In this section we present the estimated econometric models of child labor and schooling for both the older and the younger cohorts. To address the potential endogeneity of height-for-age in the schooling and child labor equations we estimate the models in two-stages the validity of which is previously discussed. In the first stage we regress the height-for-age z-scores of the children on the instruments and the other covariates in the second stage of the corresponding equation using the same sample observations. The first stage results are interesting in themselves because they show how weather shocks experienced early in life influence the subsequent physical stature of the child. In the next section, therefore, we briefly present the first stage results for the models reported later in this section.

First Stage Results

The first stage for each model estimated in two-stages is an OLS regression of the height-for-age z-scores on the relevant instruments and other covariates in the model. For the older cohort, the instruments are generated by interacting the dummy identifying the drought-affected children with three age dummies identifying those children who were at the critical stage of development at the time of the 1984 drought. The drought dummy itself is also included to see if the height-for-age was systematically different for drought affected children of all ages (not just those in the critical stage). The three age dummies are also included to see if height-for-age is systematically different for those who were at the critical stage in 1984 (not just those who were affected by drought). In addition, we include the mother's and father's height to control for the genetic variation in children's height so that the malnutrition caused by the drought explains only what is left of the natural differences in the heights of the children.

The procedure I followed in the case of the younger cohort is slightly different because of the way the rainfall shocks at the critical ages are defined. For the older cohort I essentially treated those who were age 4-6 and those who were at the critical stage but unaffected by the 1984 drought as comparison groups and those who were affected by the 1984 drought at the critical age as the treatment group. For the younger cohort as well the treatment comes from a rainfall shock experienced during the critical ages but the time at which they experienced the shock varies depending on their age and locality. Therefore, we don't have specific shock-period and age-cohort dummies to control for. The control group here as well consists of those who did not experience substantial rainfall shock during their critical years. Like the case with the older cohort we include mother's and father's height to control for the genetic variation in the heights of children. The first stage results for both cohorts are presented in table 3 below.

Table 3. The Effect of Exposure to Drought and Rainfall Fluctuations in Early Childhood on Height-for-age Z- scores (First Stage Results)

	Older Cohort			**Younger Cohort**	
	I	II	III	IV	V
	(94a&95)	(1995)	(94a&95)	(99&04)	(99&04)
Drought Affected in 1984	0.171	0.294*	0.170		
	(0.151)	(0.173)	(0.151)		
First Year in 1984	0.129	0.242	-0.058		
	(0.303)	(0.366)	(0.228)		
Second Year in 1984	0.462*	0.431	0.322		
	(0.261)	(0.311)	(0.214)		
Third Year in 1984	-0.202	-0.106	-0.294*		
	(0.199)	(0.224)	(0.171)		
Drought Affected at 1st Year	-0.094	-0.077	-0.094		
	(0.202)	(0.216)	(0.201)		
Drought Affected at 2nd Year	-0.592***	-0.487**	-0.590***		
	(0.227)	(0.246)	(0.227)		
Drought Affected at 3rd Year	0.098	-0.055	0.098		
	(0.193)	(0.207)	(0.193)		
Substantial rain def. at 1st year				0.047	0.061
				(0.223)	(0.220)
Substantial rain def. at 2nd year				-0.401**	-0.402**
				(0.179)	(0.180)
Substantial rain. def. at 3rd year				-0.243	-0.237
				(0.178)	(0.178)
Substantial rain surp. at 1st year				-0.030	-0.034
				(0.178)	(0.178)
Substantial rain surp. at 2nd year				-0.227	-0.221
				(0.180)	(0.179)
Substantial rain surp. at 3rd year				-0.094	-0.085
				(0.251)	(0.250)
Height of Mother	0.023***	0.024***	0.023***	0.028***	0.028***
	(0.007)	(0.009)	(0.007)	(0.010)	(0.010)
Height of Father	0.006	0.006	0.007	0.027***	0.027***

	(0.006)	(0.008)	(0.006)	(0.008)	(0.008)
Age	-0.078	-0.059	-0.125***	-0.020	-0.007
	(0.063)	(0.083)	(0.041)	(0.032)	(0.017)
Sex	-0.229***	-0.171**	-0.230***	-0.223*	-0.222*
	(0.075)	(0.081)	(0.074)	(0.119)	(0.118)
Household Size	-0.034	-0.023	-0.033	0.016	0.012
	(0.022)	(0.023)	(0.022)	(0.025)	(0.024)
Number of Siblings	0.053**	0.040	0.052**	-0.034	-0.030
	(0.024)	(0.025)	(0.024)	(0.035)	(0.035)
Sex of Household Head	-0.032	-0.026	-0.033	-0.102	-0.102
	(0.133)	(0.152)	(0.133)	(0.157)	(0.157)
Father's Education	0.346*	0.369**	0.338*	0.200	0.203
	(0.181)	(0.181)	(0.181)	(0.188)	(0.188)
Mother's Education	-0.060	-0.164	-0.064	-0.343	-0.340
	(0.439)	(0.491)	(0.438)	(0.260)	(0.260)
Age of Father	0.004	0.006	0.004	0.004	0.005
	(0.005)	(0.006)	(0.005)	(0.008)	(0.008)
Age of Mother	0.022***	0.020***	0.022***	0.009	0.009
	(0.006)	(0.007)	(0.006)	(0.011)	(0.011)
Agri. Land Area Owned	0.005	0.021	0.003	0.003	0.003
	(0.024)	(0.029)	(0.023)	(0.032)	(0.032)
Livestock Units Owned	0.003	-0.002	0.003	-0.011	-0.010
	(0.007)	(0.008)	(0.007)	(0.010)	(0.010)
Year=1995	-0.103				
	(0.069)				
Year=2004				0.108	
				(0.167)	
Constant	-6.894***	-7.488***	-6.365***	-11.453***	-11.535***
	(1.739)	(2.149)	(1.663)	(1.901)	(1.873)
F-stat for joint sig of instruments	2.900***	2.470***	2.910***	7.620***	7.610***
	(p=0.005)	(p=0.008)	(p=0.002)	(p=0.000)	(p=0.000)
Observations	2348	1116	2348	2241	2241

Cluster-robust standard errors in parentheses
*** p<0.01, ** p<0.05, *p<0.1

Notes: Site dummies were included in all the equations as controls for community fixed effects. Equation I represents the first stage for the multinomial logit and probit models estimated using pooled unbalanced panel data from 1994a and 1995 rounds. Equation II presents the first stage results for the bivariate probit, multinomial logit and probit models estimated using cross-sectional data from 1995 round. Equation III presents the first stage results for the random effects probit and random effects multinomial logit models estimated using unbalanced panel data from 1994a and 1995 rounds. Equation IV is the first stage for all the models estimated using pooled panel data for the younger cohort whereas equation V is the first stage for the random effects probit and random effects multinomial logit models for the same cohort. All first stage equations were separately estimated by OLS and the resulting residuals were used in the second stage equations as suggested by Terza, Basu and Rathouz (2008).

The first three columns in table 3 present the first stage results for the models estimated using data from the older cohort while the last two columns present the first stage results of the models for the younger cohort. All the first stage equations were estimated using OLS, correcting the standard errors for the household level clustering. Equations I and III are similar except that equation I includes year dummy for the survey round. Equations IV and V are also the same except that the former includes year dummy for the survey round. The year dummy is included in all the models estimated using the pooled panel data in an attempt to control for the potential confounding effects of the time-varying unobserved characteristics of the family and the child.

The set of instruments defined on the basis of exposure to the 1984 drought for the older cohort and localized rainfall shocks for the younger cohort are jointly significant in the first stage equations and generally have anticipated signs. In the results for the older cohort we have consistently negative signs for a big drought experienced during the 1st, and 2nd years while the signs for the 3rd year are mixed. For the younger cohort we have consistent negative signs for excessive rains experienced during the 1st ,2nd and 3rd years and substantial rain deficits during the 2nd and 3rd years. The sign for substantial rain deficit during the 1st year is positive but small compared to the other coefficients. In the results for the older cohort it is important to note that the drought dummy itself has in fact a positive sign implying children who belonged to the drought affected families in general had in fact bigger average height-for-age z-scores than those who belonged to the non-affected families. Therefore, isolating the effect of the drought on the group at the critical age was important for identification of its effect because that is where the negative effect is visible (as also suggested by the nutrition literature).

One remarkable observation about these first stage results is the fact that weather shocks appear to have their strongest effect on the child's physical stature during the second[16] year of his/her life for both the older and the younger cohorts. We find that both the 1984 drought and substantial localized rain shortages have large and statistically significant negative effects on the child's subsequent height-for-age when experienced during the 2nd year. According to these results being exposed to a significant drought during the second year reduces height-for-age z-scores of children by more than 0.5 points on average which represents about 25% of the mean height-for-age for this group of children. The effect of substantial rainfall deficit during the second year is somewhat similar (0.4 points or about 18% of the mean height-for-age z score for the younger cohort). Even the excessive rain has relatively larger negative effect during the 2nd year although it is not statistically significant at the conventional levels with a t-ratio of 1.26. In the case of the older cohort this large negative effect of drought during the 2nd year is observed despite the fact that second year olds in general had higher average height-for-age z-scores as demonstrated by the positive coefficient on the dummy for age-2 in 1984. This evidence supports the hypothesis that unavailability of additional food for the child will be more detrimental to the child's growth in the second 2nd year than the 1st because of the increasing inadequacy of breast-feeding as a source of nutrition for the child.

[16] Alderman et al for Zimbabwe and Glewwe and King (2001) for Philippines find similar results.

Main Results and Discussion

This section presents estimation results for the econometric models of child activity choice described in the section on econometric models and estimation methodology. Two sets of results are obtained for each of the bivariate probit, multinomial logit and probit models for child activity choices. First, each model is estimated through the standard maximum likelihood method, ignoring the potential endogeneity of the child's height-for-age. Second, all the models are re-estimated in two stages (using the procedure previously described) so as to address the potential endogeneity of our key variable of interest. In each case the standard errors are corrected for household level clustering.

For the older cohort the bivariate probit and multinomial logit models with four child activity classifications ($s_i=0, w_i=0; s_i=0, w_i=1; s_i=1, w_i=0;$ and $s_i=1, w_i=1$) are estimated using cross sectional data only from the 1995 round because the 1994a round did not collect data on child activity combinations. On the other hand, a multinomial logit model just for the main activity[17] of the child (where main activity could be 'neither work nor a student' or 'student' or 'work') is estimated using pooled panel (unbalanced) data from 1994a and 1995 rounds because child's main activity data are available in both rounds. Similarly, the probit models for child schooling are estimated using pooled panel from both rounds since data on schooling are available in both rounds. For the younger cohort all the models for child activity choices are estimated using pooled panel (unbalanced) data from the 1999 and 2004 rounds. All the models estimated with pooled

[17] Child's main activity is always 'student' if the child is attending school but for a child who is working his/her main activity is 'work' only if he/she is not combining work and schooling.

panel data include year dummies intended to control for the possible over-time variation in the un-observed child and family characteristics. In addition, the multinomial logit models for child activity choices and probit models for schooling are estimated with random effects to see if controlling for unobserved family and child heterogeneities substantially alters the results.

The multinomial logit models where the unobserved family and child heterogeneities are treated as nested random intercepts are estimated using the GLLAMM program in Stata. GLLAMM which stands for Generalized Linear Latent and Mixed Models, is conveniently designed for estimating models where the random effects can arise at multiple levels as in the case of family and individual members of the family (see a description of the model in Appendix C). However, for multi-response models like the multinomial logit and bivariate probit, the log likelihood functions do not easily converge particularly when multiple random effects are involved. With the data used for analysis in this essay most of the multinomial logit models, where the random intercepts were assumed not to vary across equations, ultimately converged but the Bivariate probit models failed. Therefore, only the multinomial logit and probit models with random effects are reported with the rest of the results.

The coefficient estimates for the bivariate probit models for child activity choices are presented in table 4 below. For the multinomial logit and probit models the coefficient estimates are presented in table A7, table A8, and table A9 in appendix A. However, joint estimation of child schooling and child labor equations as bivariate probit seems to be more appropriate as demonstrated by the highly significant correlation between the errors in the schooling and child labor equations reported in table 4. The multinomial logit model does not take these across equation correlations into account because of the IIA assumption. Therefore, my discussion of the results mainly focuses on the estimates from bivariate probit model although the key results for the other models are also reported for comparison.

Table 4. Bivariate Probit Estimates for Child Schooling and Work

	Older Cohort		Younger Cohort	
	I	II	III	IV
	Biprobit	2-Stage Biprobit	Biprobit	2-Stage Biprobit
Student				
Child's Height-for-age z-scores	0.121***	0.199	0.086***	0.251***
	(0.032)	(0.223)	(0.018)	(0.095)
Sex	0.330***	0.344***	0.284***	0.324***
	(0.082)	(0.090)	(0.066)	(0.070)
Age	0.040	0.048	0.196***	0.199***
	(0.025)	(0.034)	(0.018)	(0.018)
Agri. Land Area Owned	-0.014	-0.016	-0.000	-0.002
	(0.030)	(0.031)	(0.024)	(0.024)
Livestock Units Owned	0.011	0.011	0.023***	0.024***
	(0.012)	(0.012)	(0.008)	(0.008)
Father's Educ.-at Least Primary	0.855***	0.824***	0.400***	0.349***
	(0.176)	(0.197)	(0.104)	(0.107)
Mother's Educ.-at Least Primary	0.061	0.074	0.419***	0.472***
	(0.404)	(0.403)	(0.142)	(0.143)
Distance to Primary School	0.034	0.033	-0.027*	-0.027*
	(0.026)	(0.026)	(0.016)	(0.016)
Household Size	0.011	0.013	-0.025	-0.030*
	(0.022)	(0.022)	(0.016)	(0.016)
Number of Siblings	0.010	0.006	0.046**	0.052**
	(0.024)	(0.025)	(0.021)	(0.021)
Age of Father	-0.005	-0.005	-0.005	-0.006
	(0.006)	(0.006)	(0.005)	(0.005)
Age of Mother	0.006	0.004	0.003	0.001
	(0.007)	(0.008)	(0.006)	(0.006)
Sex of Household Head	-0.005	-0.007	0.109	0.113
	(0.136)	(0.136)	(0.095)	(0.094)
Year=2004			0.055	0.043
			(0.105)	(0.105)
Resid. from Height-for-age eqn.		-0.079		-0.169*
		(0.227)		(0.095)
Constant	-1.449**	-1.291*	-2.376***	-1.892***
	(0.568)	(0.747)	(0.266)	(0.390)
Work				
Child's Height-for-age z-scores	0.030	0.065	0.014	0.162
	(0.045)	(0.268)	(0.015)	(0.103)
Sex	-0.276***	-0.270**	-0.072	-0.035

	(0.107)	(0.117)	(0.062)	(0.065)
Age	0.032	0.035	0.138***	0.141***
	(0.032)	(0.042)	(0.018)	(0.018)
Agri. Land Area Owned	0.002	0.001	0.082**	0.081**
	(0.040)	(0.040)	(0.037)	(0.037)
Livestock Units Owned	0.012	0.012	0.001	0.002
	(0.017)	(0.017)	(0.011)	(0.011)
Father's Educ.-at Least Primary	-0.088	-0.099	-0.157*	-0.200**
	(0.241)	(0.274)	(0.090)	(0.095)
Mother's Educ.-at Least Primary	0.378	0.380	0.092	0.140
	(0.456)	(0.459)	(0.143)	(0.148)
Distance to Primary School	-0.012	-0.012	0.041**	0.041**
	(0.033)	(0.033)	(0.021)	(0.021)
Household Size	-0.043	-0.041	0.014	0.010
	(0.030)	(0.031)	(0.017)	(0.017)
Number of Siblings	-0.001	-0.002	-0.036	-0.030
	(0.030)	(0.033)	(0.023)	(0.023)
Age of Father	-0.000	-0.001	0.005	0.004
	(0.008)	(0.008)	(0.004)	(0.004)
Age of Mother	0.001	0.000	0.003	0.002
	(0.010)	(0.012)	(0.006)	(0.006)
Sex of Household Head	0.238	0.237	0.025	0.025
	(0.164)	(0.164)	(0.099)	(0.099)
Year=2004			0.093	0.080
			(0.108)	(0.109)
Resid. from Height-for-age eqn.		-0.037		-0.152
		(0.268)		(0.105)
Constant	1.207	1.273	-0.897***	-0.467
	(0.736)	(0.881)	(0.258)	(0.389)
Athrho				
Constant	-0.964***	-0.964***	-0.512***	-0.521***
	(0.101)	(0.101)	(0.053)	(0.053)
Observations	1116	1116	2241	2241

*** p<0.01, ** p<0.05, * p<0.1

Cluster-robust standard errors in parentheses

Notes: Dummies representing exposure to a big drought in 1984 at 1st, 2nd and 3rd years are used as identifying instruments for child's height-for-age in equation (II) . Dummies for substantial rain deficit and rain surplus at 1st, 2nd and 3rd years are used as instruments in (V). Mother's height and father's height were also included in all first stage equations to control for genetic variations in height. Site dummies were included in all equations to control for community fixed effects. The two-stage models are estimated using the approach suggested by Terza, Basu and Rathouz (2008) as previously discussed.

The magnitudes of the coefficient estimates for the types of non-linear models reported here are not very informative in themselves. Therefore, from the results obtained for each model estimated using pooled panel data or cross-sectional data, we have calculated the marginal effects of our key variable of interest, child's height-for-age, on activity choices at each value of the regressor keeping the values of the other covariates at their mean values. For the multinomial logit models with random effects, marginal effects are not calculated because of the lack of meaningful ways to approximate the values of the multiple random effects at which the marginal effects could be evaluated. However, the estimated coefficients for the multinomial logit models with random effects are generally very close to the corresponding coefficients obtained from the pooled data. This is particularly true for our key variable of interest, child's height-for-age (see table A7 and A8 in Appendix A). Therefore, partial effects of child's height-for-age on child activity choices obtained from the pooled data don't seem to be unreasonable. Similar issue arises in the case of random effects probit models because the values of the random intercepts are unknown here as well. But there is a common practice of computing the marginal effects for such models by assuming a zero value for the random effect of the particular case for which the marginal effect is being computed. Therefore, the marginal effects for random effects probit models obtained through this procedure are reported with the rest of the results. Following the standard practice in the literature we focus on the discussion of the marginal effects of height-for-age at mean but also present the average marginal effects and the marginal effects at sample minimum and maximum values for all the models.

The marginal effects at the mean of height-for-age obtained from all the models for the older cohort and the younger cohort are presented in table 5 and table 6 below, respectively. The average marginal effects and the marginal effects at sample minimum and maximum values are reported in tables A1-A6 in appendix A. The marginal effects of height-for-age obtained from probit models for participation in family work activities are also reported in these tables[18].

One general pattern we observe in these results is that the absolute magnitudes of the estimates are much larger in the two-stage models in all the cases perhaps implying that failing to address the endogeneity of child's height-for-age could substantially understate its estimated effect on the child's participation in schooling and family labor activities[19]. On the other hand, the two-stage estimates generally have bigger standard errors leading to lack of (or less) statistical significance for some of the marginal effects obtained from the two stage models, particularly for the older sample. Another notable pattern in these results is the general similarity in the signs and magnitudes of the estimated partial effects for the younger and the older cohort. The fact that we find generally similar results for two different cohorts is somewhat remarkable given the differences in sources of identification and the time periods at which children's heights were measured for the two cohorts.

Table 5. Marginal effects (at the mean value) of Child's Height-for-age z-scores on the Choice Probabilities of Various Child Activities (Older Cohort)

Model	p(stud=1|x)	p(stud=1, work=0|x)	p(stud=1, work=1|x)	p(stud=0, work=1|x)	p(stud=0, work=0|x)	p(work=1|x)
Biprobit	0.043***	-0.001	0.044***	-0.040***	-0.003**	0.004
	(0.012)	(0.006)	(0.011)	(0.011)	(0.002)	(0.007)
Biprobit, two-stage	0.071	-0.004	0.075	-0.065	-0.006	0.010
	(0.080)	(0.035)	(0.066)	(0.079)	(0.007)	(0.041)
Mlogit	-	-0.000	0.050***	-0.050***	-0.000	-
		(0.000)	(0.011)	(0.011)	(0.000)	
Mlogit, two-stage	-	0.001	0.052	-0.053	-0.000	-
		(0.002)	(0.076)	(0.076)	(0.000)	

[18] The coefficient estimates for these models are not reported but available from the author upon request.
[19] Alderman et al. (2001) find similar disparity between simple probit estimates and two-stage probit estimates of the effect of child's height-for-age on school enrollment in rural Pakistan.

Mlogit (Main act.)	0.041***	-	-	-0.038***	-0.003*	-
	(0.009)			(0.009)	(0.002)	
Mlogit, two-stage (Main act.)	0.091	-	-	-0.083	-0.009	-
	(0.059)			(0.058)	(0.010)	
Probit	0.041***	-	-	-	-	0.007
	(0.008)					(0.006)
Probit, two-stage	0.090**	-	-	-	-	0.007
	(0.054)					(0.041)
RE probit	0.040***	-	-	-	-	-
	(0.009)					
RE prob., two-Stage	0.064	-	-	-	-	-
	(0.058)					

***p<0.01, ** p<0.05, *p<0.1

Notes: Table 4 above and Tables A7, A8, and A9 in appendix A respectively present coefficient estimates for bivariate probit, multinomial logit, multinomial logit (Main activity), and Probit results from which these partial effects were obtained. The partial effects reported in this table were calculated at the mean value of the child's height-for-age z- scores and other regressors and the standard errors were calculated by the delta method. All models included controls for community fixed effects, child age and sex, land and livestock owned, household size and number of siblings, education of father and mother, distance to primary school, age of mother and father, and sex of household head.

The results for both the older cohort (in Table 5) and the younger cohort (in Table 6) confirm the findings in the earlier studies that access to better nutrition during early childhood enhances the child's chances of attending school later in life. This is true both in the joint models for child labor and schooling as well as the separate probit models for just child schooling (see 1st columns in tables 5 and 6). Focusing on the two-stage bivariate probit estimates, reducing the gap between the height-for-age of the sample in the older cohort and healthy American children with the same age by 1 standard deviation will increase the probability of school attendance by the former by 7.1%. For the younger cohort the corresponding estimate is 10%. Given the average height-for-age z scores of about -2, these estimates would mean that eliminating this height deficit through better nutrition and care in the early childhood would boost the chances of attending school by about 14.1% for the older cohort and by about 20% for the younger cohort. The signs and statistical significances of the estimated marginal effects of height-for-age on schooling obtained from the multinomial logit and probit models are similar to the estimates from bivariate probit models but slightly different in magnitudes. Although some of the estimated marginal effects are not statistically significant at the conventional levels of significance, it is important to note that the standard errors obtained through the delta method are generally noisy and may not be as informative[20] as the signs and magnitudes of the estimates.

Table 6. Marginal effects (at the mean value) of Child's Height-for-age z-scores on the Choice Probabilities of Various Child Activities (Younger Cohort)

Model	p(stud=1|x)	p(stud=1, work=0|x)	p(stud=1, work=1|x)	p(stud=0, work=1|x)	p(stud=0, work=0|x)	p(work=1|x)
Biprobit	0.034***	0.004	0.030***	-0.026***	-0.008***	0.004
	(0.007)	(0.004)	(0.006)	(0.007)	(0.002)	(0.005)
Biprobit, two-stage	0.100***	-0.013	0.113***	-0.064*	-0.036***	0.048
	(0.038)	(0.023)	(0.032)	(0.035)	(0.012)	(0.031)
Mlogit	-	0.012***	0.022***	-0.028***	-0.005***	-
		(0.004)	(0.006)	(0.008)	(0.001)	
Mlogit, two-stage	-	0.025	0.060	-0.061	-0.024***	-
		(0.029)	(0.038)	(0.040)	(0.007)	
Mlogit (Main act.)	0.034***	-	-	-0.029***	-0.005***	-
	(0.008)			(0.008)	(0.001)	
Mlogit, two-stage (Main act.)	0.085**	-	-	-0.062	-0.023***	-
	(0.041)			(0.040)	(0.007)	
Probit	0.034***	-	-	-	-	0.002
	(0.007)					(0.004)
Probit, two-stage	0.097**	-	-	-	-	0.043
	(0.039)					(0.031)
RE probit	0.042***	-	-	-	-	0.002
	(0.008)					(0.005)
RE prob., two-stage	0.119**	-	-	-	-	0.043
	(0.048)					(0.028)

[20] That is partly why we present the plots of the entire distributions of some of the marginal effects later in this section.

***p<0.01, ** p<0.05, *p<0.1

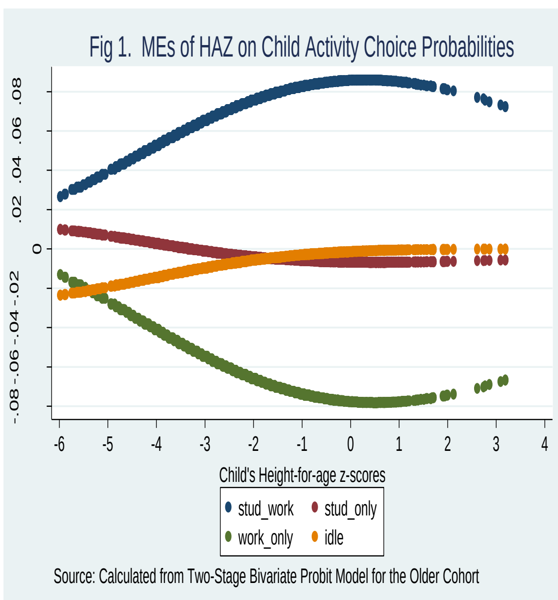
Notes: Table 4 above and Tables A7, A8, and A9 in appendix A respectively present coefficient estimates for bivariate probit, multinomial logit, multinomial logit (Main activity), and Probit results from which these partial effects were obtained. The partial effects reported in this table were calculated at the mean value of the child's height-for-age z- scores and other regressors and the standard errors were calculated by the delta method. All models included controls for community fixed effects, child age and sex, land and livestock owned, household size and number of siblings, education of father and mother, distance to primary school, age of mother and father, and sex of household head.

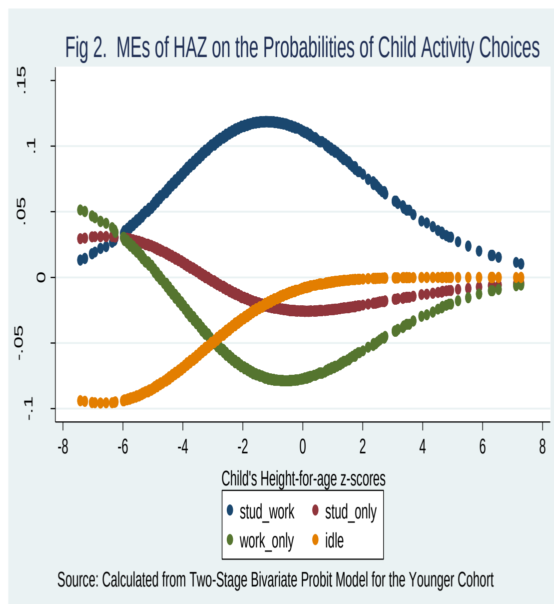
While the marginal effects at the mean of the child's height-for-age on his/her participation in family labor activities are also positive as shown along the last columns of tables 5 and 6 and the corresponding tables in appendix A, these effects are generally small in magnitude and mostly insignificant. This is so because about 89% of the children in the pooled sample for the younger cohort and 95% of the pooled sample for the older cohort were participating in family labor activities. Therefore, a more meaningful estimate would be the partial effect of the child's height-for-age on the probabilities of being selected for full-time family labor, $p(s=0, w=1|x)$. These estimates are obtained from bivariate probit and multinomial logit models and reported along the 4th columns of the tables for the marginal effects. The results show that the marginal effect of child's height-for-age on the probability of being selected for full-time family labor is consistently negative except its minimum value for the younger cohort. However, the two-stage versions of these estimates are statistically insignificant for the older cohort and mostly insignificant for the younger cohort. Based on this evidence, therefore, we find no support for the hypothesis that physically stronger children will be positively selected for full time family labor.

On the other hand, the estimates reported along the 3rd columns of the tables for the marginal effects consistently show that the physically stronger children are more likely to combine schooling and family labor than the physically less fit children. The marginal effects at mean as well as the average marginal effects of height-for-age on the probability of combining schooling and work, p(s=1,w=1|x), is consistently positive and much bigger than its marginal effects on all the other choices for child activities for both the older and the younger cohort. The results hold both in bivariate probit and multinomial logit estimates although average marginal effects and marginal effects at mean obtained from the two-stage bivariate probit models are bigger than those obtained from the two-stage multinomial logit models.

In contrast, both the marginal effects at mean and average marginal effects of height-for-age on the probability of being selected for full-time schooling , p(s=1,w=0|x), are either negative or positive but close to zero as shown along the 2^{nd} columns of the tables for the marginal effects. In addition, the marginal effects on the probability of being selected for full time schooling in the two-stage models are rarely significant while the marginal effects on combining schooling and family labor are either significant at conventional levels or generally have standard errors smaller than the estimated partial effects. Therefore, there appears to be reasonably strong and consistent evidence that better physical stature enhances the probability that the child is asked to participate in family activities while attending school but no evidence that better physical fitness increases the chances of being selected for either fulltime schooling or fulltime family labor. It is important to note that better physical fitness seems to reduce the probability that the child remains idle, p(s=0, w=0|x), although the marginal effects of height-for-age on this choice are small in magnitude particularly for the older cohort. The bottom line from these results is that, a point increase in the height-for-age z-score of the child will substantially increase the probability of combining schooling and family labor, will reduce the probability of being selected for full-time family labor, but will have little effect on the probabilities of being selected for full-time schooling or being idle.

Our discussion so far was based on the marginal effects at the mean and the average marginal effects but this may not be fully informative if the marginal effects considerably vary at different values of height-for-age. To check whether the aforementioned relationships between the marginal effects of height-for-age on various child activity choices hold at points other than the mean, we plot[21] the marginal effects against the values of height-for-age z-scores for our preferred two-stage bivariate probit model. Fig 1 and Fig 2 below present these plots for the older and younger cohorts, respectively.

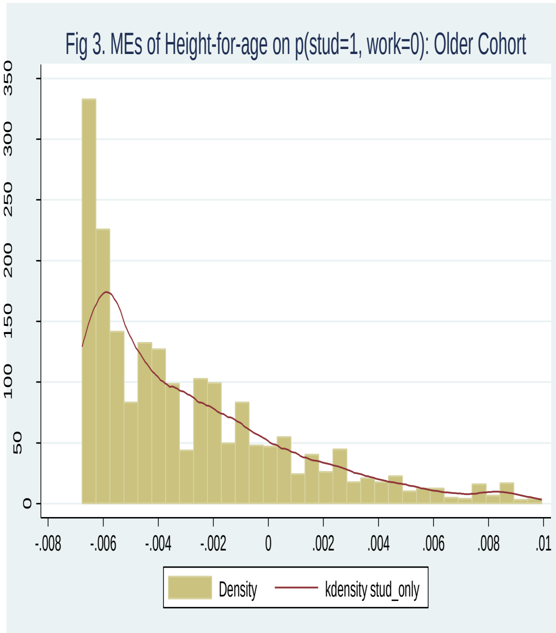


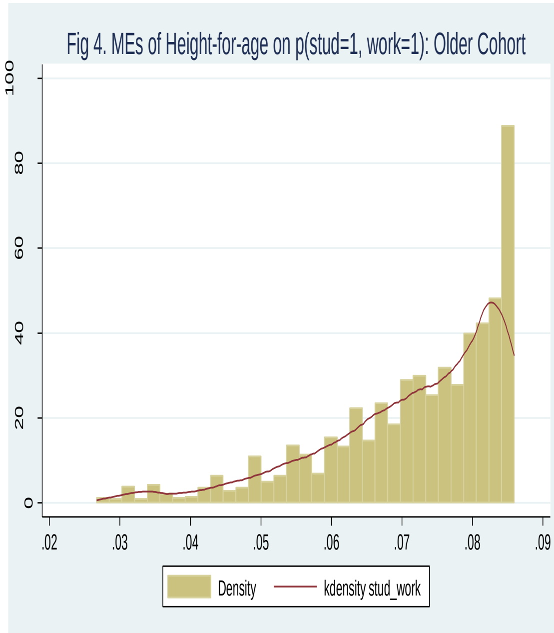


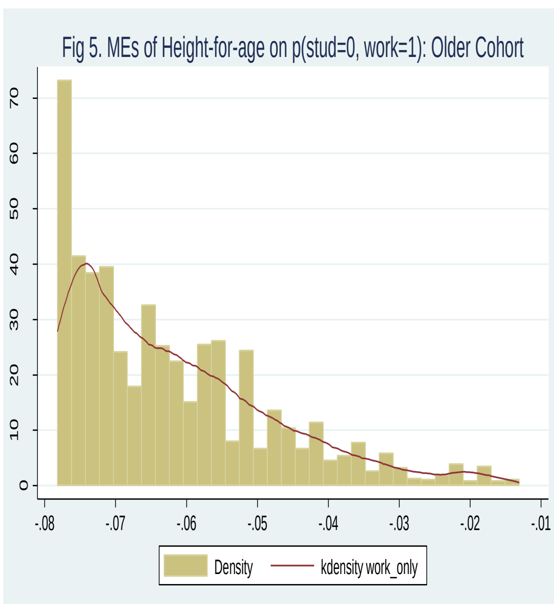
[21] While this shows how the marginal effect on the probability of each activity choice varies with changing values of height-for-age, the calculation of marginal effects at each point assumes linearity and the possible non-linearity in the effects of height-for-age on child activity choice is not addressed here. Inclusion of quadratic terms in our models doesn't seem to be informative because of the negative observations on height-for-age z-scores. An alternative way could be to estimate the models for various ranges of values for height-for-age and compare the resulting marginal effects. This is also infeasible in our case because of small sample size we are working with but future studies can address the issue using data from a larger sample.

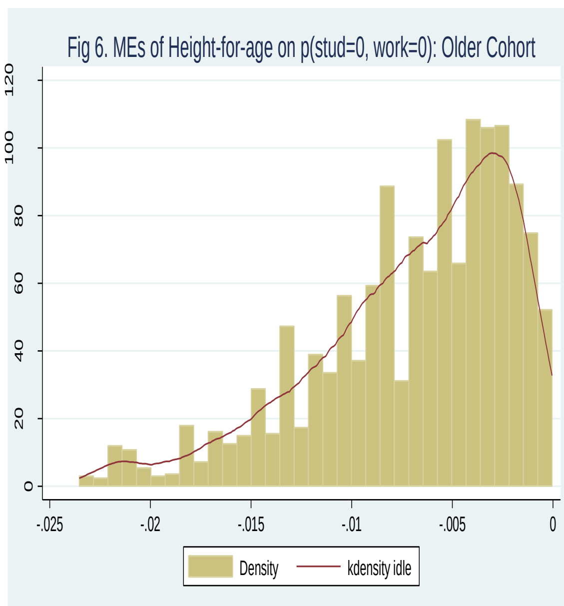
Although the observed range of values for height-for-age z-scores vary for the younger and older cohorts (-5.98 to 3.19 for the older cohort and -7.41 to 7.28 for the younger cohort), Fig 1 and Fig 2 show somewhat remarkable similarity in the patterns of the marginal effects for the comparable ranges of values of height-for-age. The marginal effects of height-for-age on the probability of combining schooling and work (stud_work) remain positive and much bigger than the marginal effects on the probabilities of being selected for other activity categories at all values of height-for-age except at the extremes. On the other hand the marginal effects on the probability of being selected for the full-time schooling (stud_only) remain close to zero for both cohorts while the marginal effects on the probability of being selected for full-time family work (work_only) remain mostly negative. The patterns in the marginal effects of height-for-age on the probability of being idle appear to differ for the two cohorts at smaller values of height-for-age but the overall pattern is similar here as well. Therefore, the relationship between the marginal effects we observed at the mean of height-for-age is not limited to that particular point but holds throughout except at the extremes where we have only a few observations and hence all the marginal effects approach zero. In fact our conclusion based on the marginal effects at the mean or the average marginal effects seems to be reasonable since most of the marginal effects are clustered around the marginal effects at the mean as demonstrated by their Epanechnikov kernel densities[22] presented in Figs 3-6 for the older cohort and Figs 7-10 for the younger cohort.

[22] The "optimal" width is used in constructing each of the kernel densities for the MEs. The optimal width is the width that would minimize the mean integrated squared error.

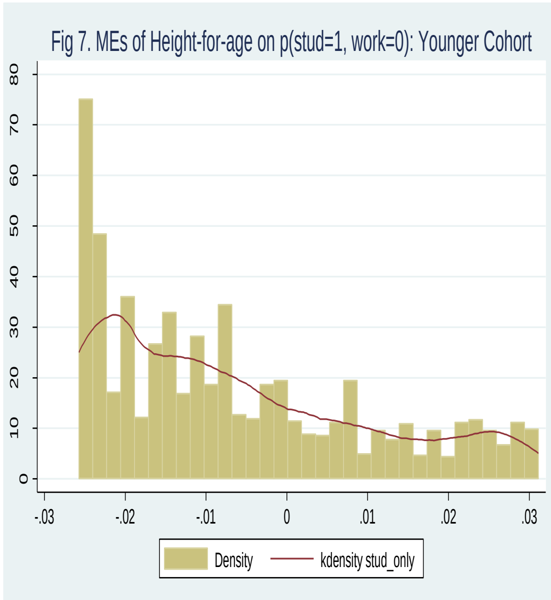


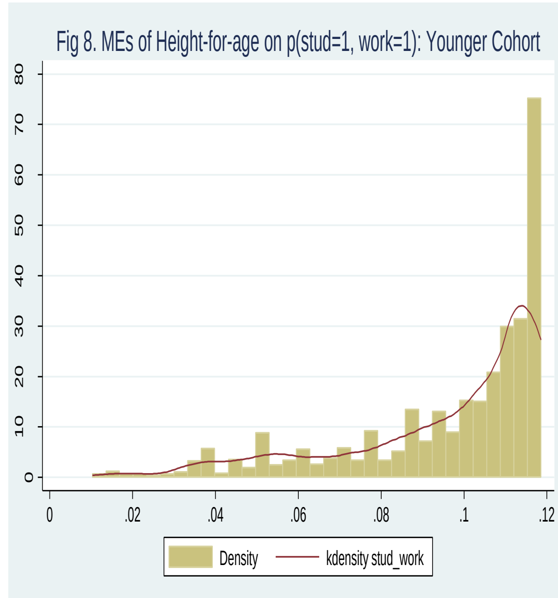


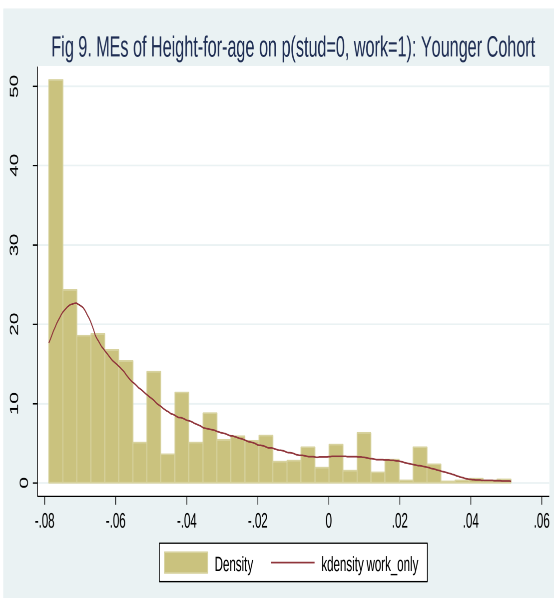


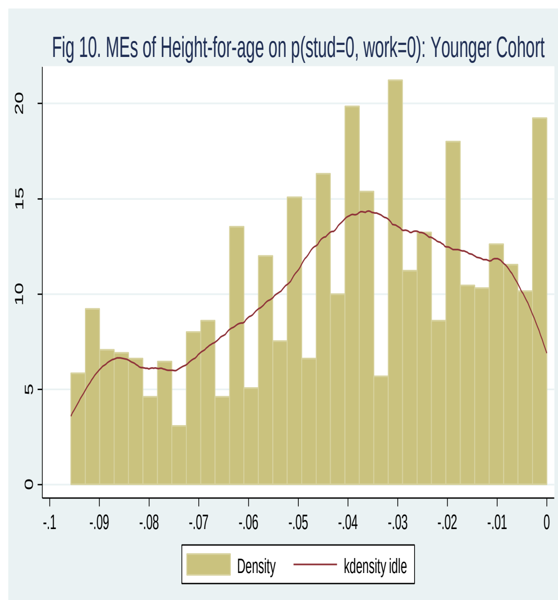


Source: Calculated from Two-Stage Bivariate Probit Model for Schooling and Work for the Older Cohort.
Notes: MEs stands for marginal effects and p stands for probability.









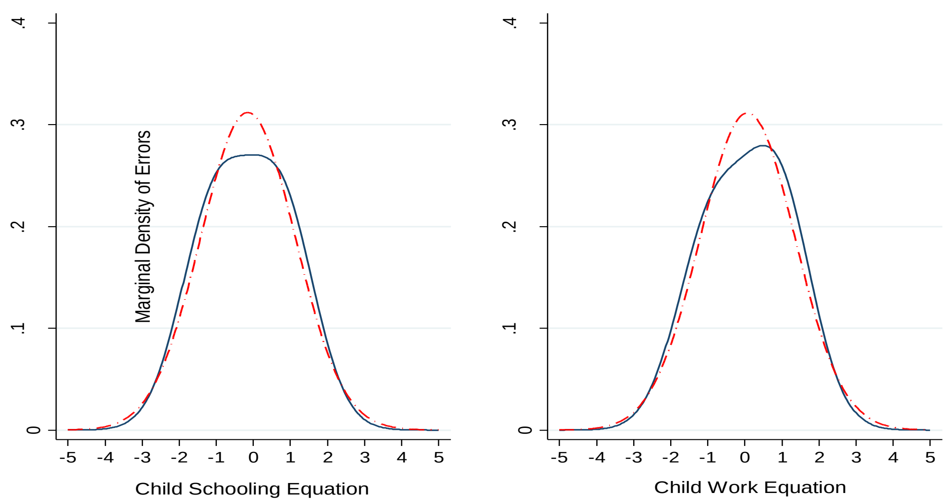
Source: Calculated from Two-Stage Bivariate Probit Model for Schooling and Work for the Younger Cohort.
Notes: MEs stands for marginal effects and p stands for probability.

With the exception of the kernel densities of marginal effects of height-for-age on the probability of being idle that are based on relatively smaller number of observations

(Fig 6 for the older cohort and Fig 10 for the younger cohort), all the other kernel densities are clearly uni-modal and highly skewed with the bulk of the marginal effects clustered around the marginal effect at the mean that itself is close to the mode of the distribution in each case. For example, the marginal effect at the mean of height-for-age on the probability of combining schooling and family work is 0.075 for the older cohort and 0.113 for the younger cohort in the two-stage bivariate probit models as shown in tables 5 and 6, respectively. The corresponding average marginal effects are 0.062 and 0.08 as shown in tables A1 and A4 in Appendix A for the older and younger cohorts, respectively. The mode of the distribution for the corresponding marginal effects is about 0.085 for the older cohort (Fig 4) and about 0.118 for the younger cohort (Fig 8) around which the bulk of the marginal effects are lumped. The same is more or less true for the marginal effects of height-for-age on the probabilities of being selected for full time schooling and full-time family labor. That is why the average marginal effects reported in tables A1 and A4 in appendix A and the marginal effects at the mean are not very far apart. Hence, the conclusions we arrived at on the basis of the marginal effects at the mean of height-for-age seem to be reasonable.

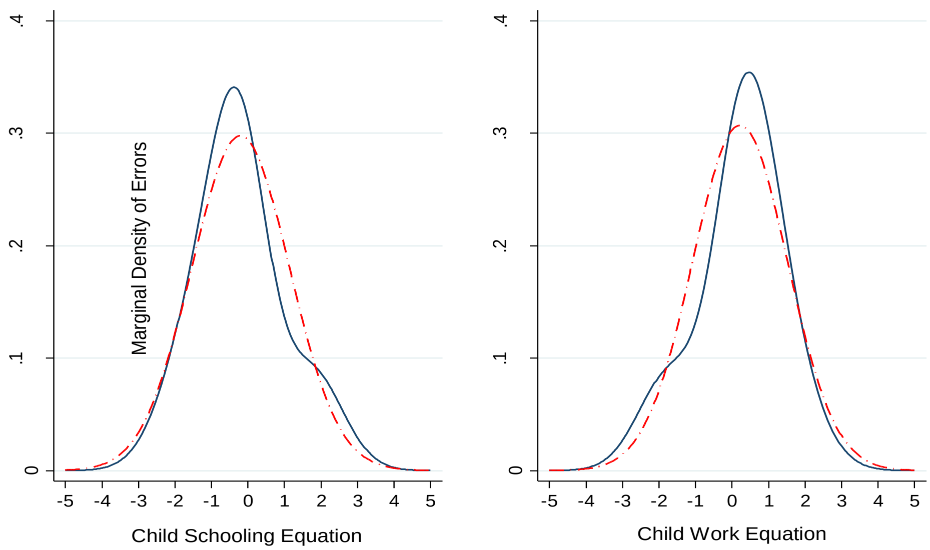
To get some feel about the validity of the bivariate probit parametric form for the joint distribution of the errors in the schooling and work equations, we tried to re-estimate the bivariate models following Gallant and Nychka’s (1987) semi-nonparametric approach previously described. Strict applications of their approach requires estimating the models for successively increasing order of the Hermite polynomial and testing the superiority of a lower order against higher order using likelihood-ratio tests or by model-selection criteria such as the Akaike information criterion or the Bayesian information criterion. With our relatively small sample of observations, however, we could hardly obtain convergence for the non-concave pseudo–log-likelihood function with Hermite polynomials of more than 2 degrees. For the older cohort the pseudo–log-likelihood function for the two stage model failed to convergence even when we set the order of the polynomial at 2 for both schooling and work equations but it converged when we set either r_1 or r_2 to 1. Fig 11 depicts the error densities from the two stage model for the older cohort when r_1=2 and r_2=1 while Fig 12 presents the error densities from the two-stage model for the younger cohort when the order of the Hermite polynomial for both equations is set to 2. The detailed characteristics of these densities along with the estimated coefficients for the covariates are presented in table A10 in appendix A.

Fig 11. Error Densities from Semi-nonparametric Estimation (2-Stage, Older Cohort)



Note: The Order of Hermite Polynomials was 2 for Schooling and 1 for Work.

Fig 12. Error Densities from Semi-nonparametric Estimation (2-Stage,Younger Cohort)



Note: The Order of Hermite Polynomials was 2 for both Schooling and Work Equations.

In the cases where convergence was obtained, the error densities are symmetric and do not seem to substantially deviate from a normal distribution (in reds) with similar first and second moments as demonstrated in Fig 11 and Fig 12 for the older and the younger cohorts, respectively. The measures of skewness and kurtosis reported in table A10 in appendix A are consistent with this observation. Although the magnitudes of the coefficient estimates from our semi-nonparametric models are not directly comparable to those from the bivariate probit model since the former depend on the order of the Hermite polynomials that we were unable to optimize, the signs and statistical significances of the coefficients are generally consistent with exception of the two-stage equation for the child work for the older cohort. Given the symmetry of the error densities and qualitative resemblance in the estimated coefficients, therefore, the bivariate probit model doesn't seem to be unreasonable for our data.

Finally, we try to check if our results for the younger cohort are being driven by the replacement of the missing monthly rainfall records with their long-term averages by re-estimating our models for the younger cohort, successively excluding the major cases with missing rainfall records from our estimation sample. The results from this exercise for our preferred bivariate probit model are reported in table B3 in appendix B. The corresponding first-stage results and marginal effects are reported in tables B4 and B5, respectively. The first adjustment we make is to limit our estimation sample to those who had at least 6 months of non-missing rainfall records including the main rainy (agricultural) season in the locality for at least 1 of the three critical years of development. The next adjustment we make is to limit our estimation sample to those who fulfill the same condition as in the first adjustment for at least 2 of the three critical years of development. And finally we limit our estimation sample to those who fulfill the same condition for all the three critical years of development. These adjustments produce little changes in the signs, magnitudes and qualities of the first-stage results, the coefficient estimates and the corresponding marginal effects as shown in the appendix. For example, we lose 283 observations because of the final adjustment but the marginal effect at the mean of height-for-age on p(stud=1, work=1) decreases from 0.113 to 0.106, its marginal effect on p(stud=1, work=0) decreases from -0.013 to -0.016, the marginal effect on p(stud=0, work=1) increases from -0.064 to -0.055 and the marginal effect on p(stud=0, work=0) increases from -0.038 to -0.036. While we still kept some cases with smaller number of missing rainfall records in our estimation sample, if there was a major understatement of our estimates because of the missing rainfall records it is likely that larger changes in the estimates would have been observed when we removed all the major cases with missing rainfall records. Hence, it doesn't seem that our results for the younger cohort are being driven by the replacement of the missing rainfall records by their long-term averages.

In general, therefore, our findings from both the older and younger cohort indicate that better access to early childhood nutrition can improve the child's prospects for attending schooling but may also put the child in additional pressure to participate in family labor activities. This may take the form of asking the child to miss classes in order to help with family labor activities at home, for example, when agricultural activities are at their peak during harvesting season or the child may be asked to look after the livestock or fetch drinking water or fuel wood after coming back from school or during the weekends. While the data at hand do not contain information on the child's performance at school and hence do not allow analysis of how performance may be affected by the child's physical stature, it is quite possible that the additional pressure put on the child's time from the family labor activities could constrain the amount of time the child could spend on home works and other school related activities at home and hence lead to poor performance at school. Therefore, policies that try to promote schooling through nutrition support programs could be more effective if they are accompanied by programs that could mitigate the forces that push families to resort to child labor.

Conclusion

This essay examines how malnutrition experienced during early childhood influences the subsequent participation of the child in schooling and child labor activities in rural Ethiopia. The cumulative outcome of the child's early childhood nutritional experience is measured by the child's height-for-age z-score which is also taken as a measure of the child's physical fitness. Our theoretical model implies that the effect of the child's physical fitness on the parental choice as to whether to select the child for schooling or child labor is ambiguous. Bivariate probit and multinomial logit as well as separate probit models are estimated to empirically examine the effect of the child's physical fitness on his/her participation in schooling and family labor activities. Data from various rounds of a unique longitudinal rural household survey in Ethiopia (Ethiopian Rural Household Survey) are used to estimate the models. To address the potential endogeneity of the child's physical fitness in the models for child activity choices, we estimate the models in two-stages. Exposure to a famine caused by a massive drought in 1984 is used as an exogenous source of early childhood malnutrition for the older cohort of children for whom the models were estimated. Localized rainfall shocks were used as a source of identification for the younger cohort of children for whom a separate set of estimates were obtained.

The first stage results show that exposure to significant weather shocks during the first three years of the child's life generally have a lasting negative effect on his/her age standardized heights measured later in life. The effect is particularly strong when the child is exposed to the shock during his/her second year. Estimation results from the child's activity choice models indicate that better early childhood nutrition enhances the child's chances of attending school later in the child's life. The range of marginal effects obtained from the two stage bivariate probit models imply that equalizing the median height of the children in the sample for the two cohorts with the heights of healthy

American children of the same age through better nutrition and care in the early childhood would boost the chances of school attendance among the children by at least 14% and possibly by as much as 20%.

On the other hand, we find no conclusive evidence that better physical fitness of the child leads to his/her positive selection for full-time child labor activities. We rather found reasonably strong and consistent evidence that physically robust children are more likely to combine child labor and schooling than physically weaker children. The results are consistent across two different cohorts of children and two different identification strategies. The findings indicate that, although better early childhood nutrition leads to higher chances of attending school, it may also put the child at additional pressure to participate in family labor activities and this may be reflected in poor performance in schooling. Therefore, policies that try to promote schooling through nutrition support programs could be more effective if they are accompanied by programs that could mitigate the forces that push families to resort to child labor. My next work in this area will look at how the observed effect of physical fitness on the probability of combining child labor and schooling affects the child's school performance in the form of test scores and grades.

ESSAY II: MOTHER'S BARGAINING POWER, GENDER RATIO PREFERENCES AND CHILDREN'S HUMAN CAPITAL OUTCOMES: EVIDENCE FROM SIBLINGS AND TWINS

Introduction

A lot of theoretical and empirical research has been conducted to explain the influence of family demographics and resources on investment in human capital of children. The key theoretical development in this area has been the formulation of the quantity-quality trade off model of fertility by Becker (1960) and others. The model is based on the premise that parents are inclined towards equalizing the quality among their children (Becker and Lewis 1973; Becker and Tomes 1976) and it explains the interdependence between the shadow prices of quantity and quality per child. Becker's model has motivated a large number of empirical studies on the effects of family size on educational outcomes of children. The roles of child's gender, birth order and child age gaps have also been widely investigated.

The Becker model, however, did not address how differences in parental preferences for gender mix of children and decision-making power in the family could influence the distribution of quality across children. This is so because Becker's model essentially assumed a 'unitary' household where mother and father are treated as having a joint utility function. Other theoretical models like the family 'bargaining model' (McElroy and Horney 1981; McElroy 1990) and 'collective' household decision making model (Chiappori 1988, 1992) were later developed to explain the implications of differences in maternal and paternal preferences and bargaining power. The implications of differential bargaining power between the mother and the father for fertility and the distribution of quality among boys and girls have been examined by some empirical studies as well (see the literature review section). While the role of explicit parental (maternal or paternal) preference for gender mix of children in influencing the distribution of quality across children has not been directly addressed in the existing literature, a related issue of the role of siblings' gender composition on child health and education has been lately investigated by some empirical studies (Butcher and Case 1994; Hauser and Kuo 1998; Kaestner 1997; Morduch 2000; Garg and Morduch 1998; Paris and Willis 1993). The findings from these studies are mixed and vary from country to country perhaps, reflecting differences in the cultural values that determine parents' attitudes towards human capital for male versus female children as well as the type of resource constraints that they face. In addition, none of the studies conducted so far provides conclusive evidence as to the causal mechanism that drives the relationship. In particular, the role played by the explicit parental preference for gender mix of children (combined with the parents' bargaining power in the family) in shaping the distribution of human capital across children has not been investigated.

This essay attempts to expand the literature in this area and contribute towards better understanding of the nature of the relationship between siblings' gender structure and parental investment in their health and education using data from siblings and twins in a sample of Ethiopian (and Indian) households. Specifically, an attempt is made to examine the extent to which the interaction between a mother's revealed preferences for the gender mix of her children (vis-à-vis the actual gender mix) and her bargaining power in the household could lead to differences in distribution of education and health among her children. The key question addressed here is whether the mother uses her bargaining power in the family, say in terms of controlling the distribution of nutrition inputs, to influence the distribution of child quality in the direction of her gender ratio preferences when the actual ratio deviates from her preferred ratio. For example, if the mother prefers to have equal number of boys and girls but ends up with 75% boys and 25% girls, does she try to compensate for the deficit in the quantity of girls by more heavily investing in the quality of girls given her influence on resource allocation in the household? This of course depends on why the mother has a preference for a child of one gender over another to begin with. For example, if the mother wants to have more girls in order to help her with family activities at home, we expect one girl born among many boys to have less chance to be sent to school than one among many girls. On the other hand, if the mother's preference for girls is a reflection of her wish to have daughters that are successful in life, then a girl born among many boys is likely to have a better chance to go to school than one born among many girls. The same reasoning applies to a boy's chance of attending school.

The key point of departure for this study is that a mother may have distinct preferences not only for the number and quality of children but also for the gender mix of her children that may not be consistent with the father's preferences. Theoretically, all the three are choice variables that could be influenced by the actions of the mother, although a parent's ability to determine gender mix is limited. The mother could control the number of children through contraception and the gender ratio through selective abortion or, in the extreme case, by neglecting a child of the unwanted gender to death. In an underdeveloped country like Ethiopia, however, the technological options for abortion are quite limited and there is no documented evidence of extreme bias against girls or boys to the extent that they are neglected to death. Consequently, the gender ratio of children is largely beyond the control of the mother and hence there is likely to be a gap between a mother's preferred and actual gender ratio. But the mother's actions can influence the quality mix of the children, the degree of the influence depending on the mother's bargaining power in the household. So, an important question is whether mothers use their bargaining power to influence the distribution of quality among their children so as to compensate for the imbalance between the desired and actual quantity of children. This essay attempts to empirically examine this question using data from twins and siblings from the demographic and health surveys of Ethiopia with robustness checks using similar data from India.

Literature

The question addressed in this essay is related to at least two key issues that have lately attracted increasing attention from researchers in economics and other social sciences. The first is the role of mother's power in influencing the distribution of well-being across children and the second is the effect of siblings' gender composition on the human capital of a child. There are several indicators of women's bargaining power in the family that have been used to analyze the effects of specific aspects of mother's power. Kishor (2000) classifies these measures into three categories – direct evidence of women's power, indicators of sources of empowerment and indicators of the setting for empowerment. The first represents the final outcome while the other two represent the processes that lead to the final outcome. The indicators of bargaining power that are often used in economics literature, notably employment, education, non-labor income and transfers, assets brought into marriage, income shares of women, inheritance, current asset shares, age at marriage and media exposure, fall under what Kishor calls sources of empowerment. These typically represent the potential for but do not guarantee mother's empowerment. Other measures that are often used, for example the education gap between husband and wife, literacy gap, age gap and past incidence of violent treatment (beating) by husband, fall under settings for empowerment since these largely refer to the circumstances of the current and past life of the mother. Direct evidence of women's power includes such factors as control over expenditures and purchases, freedom of movement, attitudes towards violent treatment from husband and control over fertility choices. One or more of these indicators have been separately used to analyze the effects of specific aspects of mother's power.

The emerging consensus regarding the effect of mother's empowerment appears to be that mother's control over resources and decision making process in the family enhances the welfare of children. While this generalization is largely based on studies conducted in developing countries, a few studies that used data from developed countries also arrived at similar conclusions. For example, Lundberg, Pollak, and Wales (1997) exploited a natural experiment generated by the replacement of a big part of child tax allowance by direct cash payment to the mother under the UK Child Benefit scheme of 1977 and found that the transfer significantly increased the fraction of budget spent on clothing for women and their children. Koenen, Lincoln and Appleton (2006) also found women's power measured in terms of political participation, economic autonomy, employment, earnings, and reproductive rights at the state level in the United States to be associated with better child-well-being measured in terms of child birth weight, infant and teen mortality, high school dropout rate and teen birth rate.

There is a relatively larger number of studies conducted on the effect of women's power on child well-being in developing countries mostly focusing on child nutrition and health as a measure of well-being. The most commonly used measures of women's power include non-labor-income, education, employment and asset ownership, often measured relative to the husband's power. Some direct indicators of control over family resources and decision-making processes as well as cultural and demographic variables that may influence mother's relative power (such as religion and age) have also been used. For example Thomas (1990) and Schultz (1990) found that non-labor income received by mothers is more likely to be spent on child health than income received by fathers. Smith and Haddad (2000) used the female-to-male life expectancy ratio as an indicator of women's status across 63 developing countries for the period from 1970–96 and found that greater women's status reduces the percentage of children who are underweight. Thomas (1994, 1997), Glewwe (1999), Desai and Alava (1998) all used relative education as a measure of mother's power and found a positive relationship between mother's empowerment and some measure of child health, although the relationship is not conclusive in the study by Desai and Alava after controlling for other relevant variables. Thomas (1997) used household and individual level data from Brazil and found women's income to have a significant and larger positive impact on child nutritional status than does similar income accruing to men.

Some empirical studies have also shown that higher mother's bargaining power leads to more child schooling (e.g., Thomas 1990; Schultz 1990; Adato et al. 2003).

However, in a theoretical model Basu (2006) demonstrates conditions under which too much bargaining power for the mother relative to the father could result in a decline in school enrollment when children are needed for child labor as well. In support of Basu's model, Gitter and Barham (2008) used conditional cash transfer programs under a Nicaraguan social safety net program as a randomized experiment and found the ratio of mother's education to that of the father to have a positive effect on child schooling up to some level, but a negative effect in the case of too much power imbalance in favor of the mother. Using data from India, Lancaster, Maitra, and Ray (2006) provide similar empirical support for Basu's model. Therefore, the empirical evidence in support of the positive association between mother's empowerment and child schooling does not appear to be as strong as the evidence in support of its effect on child nutrition.

Some of the empirical studies specifically examined whether increasing women's empowerment has a differential impact on health and schooling of boys and girls. For example, Thomas (1994) examined the effect of relative parental education and non-labor income on child height in the United States, Brazil and Ghana and found that increasing mother's power favors daughters over sons while father's power favors sons over daughters. Thomas (1990) also found evidence that mother's power favors the nutrition of daughters while father's power favors sons. Duflo (2003) examined the effect of the expansion of the benefits and coverage of the South African social pension program to the black population in the early 1990's and found positive and significant effects on girls'(granddaughter's) weight-for-height when the recipient is a woman (grandmother), but little effect on boys. Neither girls' nor boys' well-being was significantly affected when the recipient was a man (grandfather). Emerson and Souza (2007) find that in

Brazil mothers' education has a greater impact on education of daughters than sons' while the opposite is true for the father's education. One other study by Gibson (2008) examined the effect of absence of fathers on the survival and growth of boys and girls in one region of Ethiopia and found that girls benefit more than boys in terms of both indicators. However, there are a few contrasting studies, particularly for Africa, that find mother's power favoring sons and father's power favoring daughters (Haddad and Hoddinott 1994; Quisumbing and Maluccio 2003). The latter find that in Ethiopia, mother's assets at marriage decrease girls' chances of education while father's assets increase their chances.

While the literature shows that there is some evidence supporting the claim that mother's power may have gender specific effects on the distribution of nutrition and schooling across children, the causes for such relationships are not well understood. For example if the mother's empowerment ends up favoring daughters over sons, is it because the mother provides preferential treatment to her daughters over her sons or just because the mother's power creates a family environment that daughters can exploit to their advantage (like learning from their educated mother) even if the mother did not intend to intentionally discriminate among her children? Do such observations persist in the situations where mothers prefer to have more sons than daughters? How does the gender composition of the children the mother actually has (vis-à-vis what she would prefer to have) influence her role in distributing education and nutrition across the children? These questions, while interesting and warranting answers, have not been explicitly investigated in the existing literature.

Studies that examined the effect of gender composition of siblings on education and health of a child have not explicitly addressed the possible role that could be played by the differences in the mother's and father's preferences for gender mix of children in the subsequent distribution of quality across children. Neither did they address the role of the bargaining power of the mother or the father in influencing such distributions in the context of explicit preference for gender mix of children. For example Butcher and Case (1994) examined the effect of having sisters vs. brothers on women's educational attainment using three different data sets for the U.S (PSID, NLSY, and CPS) and found out that having sisters diminishes women's educational achievements. On the other hand, Conley (2000) finds that the number of opposite sex siblings hurt educational attainment of a given child. But they have not explicitly investigated how this could be related to parental preferences for gender mix of children and distribution of bargaining power within the family.

A few studies that tried to investigate similar issues in developing countries (Morduch 2000; Garg and Morduch 1998; Paris and Willis 1993) also found some evidence that the gender mix of the siblings can influence the human capital outcomes of a child. However, unlike Butcher and Case (1994), which found a negative relationship between having sisters and educational attainment of women for the U.S, Garg and Morduch (1998) find positive correlation between number of sisters and nutritional health outcomes for both boys and girls in a sample from Ghana. Morduch (2000) also found some weak evidence that the number of sisters may influence the child's education in Tanzanian data but finds no evidence as such in the South African data.

Prior literature thus shows that the nature of the relationship between siblings' gender composition and human capital outcomes of children is so far inconclusive. The relationship may be a weak one or may vary from country to country, perhaps depending on the cultural values that determine parents' attitude towards the gender distribution of children's human capital as well as the type of resource constraints that parents face. In addition, none of the studies conducted so far provides conclusive evidence as to the causal mechanism that drives the relationship. In particular, the role played by the explicit parental preference for gender mix of children on the distribution of human capital has not been investigated. This study tries to expand the literature in this area by specifically analyzing the role played by the mother's preference for gender mix of her children and her bargaining power in the family in influencing the distribution of health and schooling investments across her children.

Theoretical Background

Becker's original model of fertility, which analyzed the tradeoff between quantity and quality of children, assumed a 'unitary household' where husband and wife are treated as a single decision-maker with common preferences. Such a model does not accommodate the possible differences in preferences of the mother and father towards the quantity and quality of children and other aspects of fertility. Household models that allow for intra-household differences in preferences were developed by Becker (1991), McElroy and Horney (1981), McElroy (1990), and Chiappori (1988, 1992), among others. Household bargaining models as in McElroy and Horney (1981) define what are called ‘outside options’ or ‘fallback positions’ for each partner and treat the household decision problem as a bargaining game leading to Nash Equilibrium outcomes. However, the equilibrium outcomes in such models largely depend on the definition of the outside options or ‘threat points’ and hence may not provide predictions about the behavior of the partners that could easily be tested with the commonly available data (Strauss and Thomas, 1995). For example the bargaining model was criticized by Ulph (1988), as cited in Kanbur and Haddad (1994), for failing to predict choice behavior in the event of a breakdown in bargaining. On the other hand, the collective models of household behavior as in Chiappori (1988, 1992) explicitly assume that collective decisions by the partners will reflect Pareto efficiency in intra-household allocation, and thus could be more restrictive but provide more readily testable implications about the choices that the partners make without making any specific assumptions about individual preferences, except that they can be represented by a well-behaved utility function. As a result the latter approach is followed here to describe the household's decision problem involving the allocation of health and education investments among boys and girls where the mother and father may have different preferences.

For a household consisting of a mother, a father, and n children (m daughters and k sons) a general utility function (U) may be defined as a weighted sum of the mother's and the father's separate utility functions where the weights reflect the bargaining power of the mother and the father as in Chiappori and Browning (1998) or Basu (2006). Letting θ[23] $\in [0,1]$ and ($1-\theta$) represent the bargaining power of the mother and the father

[23] Typically θ will be a function of some choice variables in the model like hours worked and income as argued for example by Basu (2006) but for the theoretical exposition here θ is assumed to depend only on exogenous variables beyond the control of the household. For the empirical analysis, some self reported

respectively, the household's utility function defined over consumption of goods and leisure for every member, and health and education for the children can be stated as,

$$U = \theta u^m(c, l^m, l^f, l_1, ..., l_n, s_1, ..., s_n, h_1, ..., h_n; n, r, \gamma^m, \eta) + (1-\theta)u^f(c, l^m, l^f, l_1, ..., l_n, s_1, ..., s_n, h_1, ..., h_n; n, r, \gamma^f, \eta) \quad , \qquad (2.1)$$

where c is total consumption by all the household members, l^m, and γ^m are mother's leisure and a vector of individual and household characteristics that may affect mother's utility, respectively; l^f and γ^f are father's leisure and a vector of individual and household characteristics that may affect father's utility, respectively; l_i ,h_i and s_i are child i's leisure, health[24] and time spent in schooling, respectively; n is the number of children, r is the ratio of girls to total children, and η is a vector of household characteristics that may affect household welfare. The number of children and gender ratio are assumed to be non-choice variables which are largely true in a poor country like Ethiopia where the means for birth control and selective abortion are beyond the reach of most of the families. If the current wage rates of the mother, father, and child are given by w^m, w^f, and w_c respectively and everybody has total time endowment of 1, the full income budget constraint for this household can be stated as,

$$c + \sum_{i=1}^{n} p_h h_i + \sum_{i=1}^{n} p_s s_i \le v^m + v^f + (1-l^m)w^m + (1-l^f)w^f + \sum_{i=1}^{n}(1-l_i - s_i)w_c \,, \qquad (2.2)$$

direct measures of mother's empowerment and predetermined sources of mother's power like education gap are used as indicators of bargaining power.

[24] The treatment of child health as a variable fully controlled by parents is a bit simplistic. Parents' actions will be part of the inputs into the health production function that generates child health which will also depend on the environment and child health endowments as in Thomas (1994) for example. For simplicity, however, the model assumes that parents can choose specific level of child health.

where, v^m and v^f are the non-labor incomes for the mother and father, respectively; p_h and p_e are the direct costs of health and education per child, respectively. The prices of consumption goods are normalized to 1. For simplicity, the direct costs of health and education are assumed to be the same for boys and girls. The wage rate is also assumed to be the same across boys and girls. Where the outside market for child labor does not exist, w_c would be the implicit price for the time spent in child labor. The model stated here also assumes that children do not have direct influence on the decision making process in the household and they do not earn non-labor income.

The household's problem is to maximize the total utility (welfare) function defined under (2.1) subject to the full income budget constraint stated under (2.2). Assuming that the utility functions of the mother and father are well-behaved (strongly concave, twice differentiable and increasing in c, l^m l^f, l_i, s_i, and h_i), that the non-negativity constraint for consumption is non-binding, that the time constraint for both parents and child is non-binding and that investment by the parents in schooling and health of a child are independent[25] decisions, we can maximize (2.1) subject to (2.2) to obtain the reduced form demand[26] functions for child i's health and schooling as a function of predetermined variables as,

$$h_i = f_i(v^m, v^f, w^m, w^f, w_c, p_h, \gamma^m, \gamma^f, \theta, r, n, \lambda) \tag{2.3}$$

$$s_i = f_i(v^m, v^f, w^m, w^f, w_c, p_s, \gamma^m, \gamma^f, \theta, r, n, \lambda) \tag{2.4}$$

---

[25] The first essay in this dissertation specifically deals with the interdependence between health, schooling and child labor decisions by the parents.

[26] Detailed solutions and characterizations of the demand functions that come out of the collective model of household behavior are presented in Chiappori (1988, 1990), Browning and Chiappori (1998) and Basu (2006) among others.

where λ is the multiplier for the full-time budget constraint stated under 2.2. The key variables of interest for this study are the proportion of girls among the total children (r), the mother's bargaining power in the family (θ) and the number of children (n). The separate effects of each of these variables on child health and education have been investigated by a number of other studies. In the absence of parental preferences for gender mix of children (r), intra-household resource allocation models predict a tradeoff between the number of children (n) and health and/or education that each child receives as demonstrated in Becker's model of fertility. This is based on the assumption that parents want approximately equal distribution of quality across their children and resource constraints that families face would imply that the larger the number of children the smaller the quality of each child will be. Therefore, families who want higher education and better health for each of their children will decide to have a small number of children and families who prefer to have large number of children will have to bear lower health and education for each child.

When one or both parents have specific preferences for the gender mix of their children (r), however, new implications emerge that are not necessarily consistent with the predictions of the standard models of fertility and investment in child quality. First, it is possible that parents who fail to achieve the desired gender mix during the initial births may keep trying in the hope that the new trials produce the desired gender mix[27]. As a result, the prediction of the Becker model that parents limit the quantity of children so as to improve quality of each child may not hold; i.e., quantity may increase and quality may be compromised in the pursuit of the desired gender mix of children. Second,

[27] There is some empirical evidence that supports the claim that families with same-sex siblings tend to be larger (for example, Angrist and Evans 1998; Baez 2008).

parents may also attempt to compensate for the deficit in the quantity of children of the preferred gender by more heavily investing in their quality, henceforth called the compensating hypothesis. Therefore, the premise upon which the Becker's model is based, that parents want to equalize quality among their children, does not necessarily hold. The validity of this implication; i.e., whether parents try to compensate for the deficit in the quantity of children of the preferred gender by more heavily investing in their quality, has not been empirically examined. It will be interesting to look at whether the mother (who the literature says has a bigger role in child-well-being than the father) influences the distribution of investment in child quality in the direction of her gender preference when the actual gender mix of kids deviates from her preferred mix.

Whether the compensating hypothesis holds or not in the case of child schooling will depend on why the mother has a preference for a child of one gender over another to begin with. For example, if the mother wants to have more girls just because she wants them to help her with family activities at home, it is likely that a girl born among boys will have less chance to be sent to school than one born among some boys and some girls. On the other hand, if the mother's preference for girls is a reflection of her wish to have daughters that are successful in their future life, then a girl born among boys is likely to have a better chance to go to school than one born among some boys and some girls and hence the compensating hypothesis will hold. Similar argument applies in the case of preference for boys.

For the mother to be able to influence the distribution of child schooling and health one way or another, however, she must have some degree of control over the decision making process in the family. If the mother's preference for girls/boys is a reflection of her wish to have daughters/sons that are successful in their future life, not just to help her in family labor activities, the mother's bargaining power will reinforce the distribution of both child health and schooling in the direction implied by the compensating hypothesis. It is the effect of this interaction between mother's gender preference (vis-à-vis actual gender mix of her kids) and bargaining power on the distribution of child health and education on which the empirical analysis in this essay focuses.

Econometric Model

The key task in this essay is to examine the effect of the disparity between the gender ratio preferred by the mother and the actual gender ratio of her living children on health and schooling of a child. The extent to which the mother can influence the distribution of education and health among her children based on her gender preferences, however, may depend on her bargaining power in the household. Therefore, the empirical models for child health and schooling include the interactions between the mother's gender ratio gap (actual ratio less desired ratio) and a measure of the mother's bargaining power (θ) as regressors.

As previously discussed, there are several indicators of women's bargaining power in the family that have been used to analyze the effects of specific aspects of mother's power on child health and schooling. For the analysis here measures of mother's bargaining power are constructed from multiple indicators using factor analysis[28]. This approach is particularly appealing for the purpose at hand because there is a general

[28] A similar approach was followed by Kishor (2000), Jejeebhoy and Sather (2001), Smith et.al (2003), Varadharajan (2003), and Ahmed (2006), among others.

consensus in the literature that women's status in the family is a multidimensional concept that may not be effectively represented by one proxy as is often done in the economics literature in this area.

Factor analysis helps us to condense a large number of variables measuring multiple dimensions of a given concept (like women's power) into a smaller number of non-overlapping measures that summarize the common variation in the original variables. For example, if we have j observable indicators of bargaining power $x_1,....x_j$, for n mothers, the variability in these related indicators of mother's bargaining power can be expressed as a linear combination of $q \leq j$ unobservable factors and residuals as stated, for example, in Velicer and Jackson (1990),

$$x = Lf + e \tag{2.5}$$

where, x is j X n matrix of observable indicators of mother's bargaining power, f is q X n matrix of factors, L is j X q matrix of factor loadings representing the correlation coefficients between j indicators of bargaining power and q factors and e is j X n vector of errors that are uncorrelated with f and assumed to be independent of each other. The elements of the factor loadings matrix can be obtained through regression. The squared factor loading is the percent of variance in that indicator variable explained by the factor and the eigenvalue for a given factor measures the variation in all the variables which is accounted for by that factor.

One problem with factor analysis is that there is no definite rule to determine the number of factors that sufficiently capture the variation in the original data. The most commonly used criteria in determining the optimal number of factors to be extracted are Kaiser-Guttman rule and the Cattell's scree test (Hayton, Allen and Scarpello 2004). The

Kaiser-Guttman[29] rule states that only those factors with eigenvalue greater than 1 should be retained. The scree test involves an examination of a plot of the eigenvalues for breaks or discontinuities. The criterion for retention is to keep those factors with eigenvalue exceeding the point at which the sharp fall in the plot of the eigenvalues stops. In our application we examine both the magnitudes of the eigenvalues and their scree plots to determine the number of factors to retain.

The indicators of mother's power that are used in the factor analysis to construct condensed measure(s) of empowerment are education or literacy gap between husband and wife, age gap, media exposure (frequency of listening to radio), presence of other wives for the same husband, and multiple self-reported indicators of freedom from violence from husband (the conditions under which the wife thinks beating by husband is justified). Information on some other commonly used measures like non-labor income and separate asset ownership are not available in the data used for analysis in this essay. Mother's employment status, which could be both a cause and an outcome of the mother's bargaining power, is not included in the indicators of bargaining power because of its apparent endogeneity[30] in the child schooling and health equations.

The measure(s) of mother's power so obtained is included in the empirical models for child health and education as a regressor along with its interactions with the gap between the desired and actual gender ratios of her living children. The desired gender ratio is defined as the ideal number of girls the mother would like to have divided by the

---

[29] The theoretical basis for this criterion is demonstrated in Guttman (1954).

[30] For example, employment may provide a mother with higher bargaining power but may also leave her with limited time to spend on childcare, thus directly affecting the health of the child. Therefore, unobserved family characteristics that may be correlated with mother's opportunities for employment may also be correlated with child's health and schooling.

ideal number of boys and girls she would like to have. The actual gender ratio is defined as the ratio of total number of living girls for the mother to the total number of living children. By focusing on the gender ratios instead of absolute number of girls or boys, we are assuming that the mother tries to allocate the health and education investment across boys and girls at each point in time expecting the next child to be a girl or a boy with equal probability and hence anticipating the realized gender ratio at that point in time to remain the same, on average. That is, the mother thinks that whatever lopsidedness observed in the gender mix of her children thus far is an anomaly and that it need not repeat itself if she decides to have more children.

The dependent variable in the demand for child health equation is the child's weight-for-height z-score that is commonly used as a short-term measure of the child's physical fitness. The mother's influence on the child's weight-for-height is presumed to come mainly through her control over the distribution of food in the household although she may also influence other aspects of childcare at home and outside that may be reflected in his/her physical fitness. Schooling is represented by current enrollment status. Assuming that the parents' demand for child health and education are linear functions of family and child characteristics, the family's demand for child health or school enrollment (as a linear probability model[31]) could be stated as,

$$\begin{aligned} y_{ij} &= \delta_0 + \delta_1 S_{ij} + \delta_2 \theta_j + \delta_3 (r_j - d_j) + \delta_4 (r_j - d_j) * S_{ij} + \delta_5 (r_j - d_j) * \theta_j + \\ &+ \delta_6 (r_j - d_j) * \theta_j * S_{ij} + X_{ij}^{'} \beta + Z_j^{'} \gamma + \mu_j + \varepsilon_{ij} \end{aligned} \tag{2.6}$$

[31] The left hand side for a linear probability model for school enrollment is *Pr(y_{ij}=1|.)*. We avoided probit model for this dichotomous outcome variable because implementation techniques for interactive instruments are not yet available in such non-linear framework. However, linear approximations of such models provide consistent estimates of the average treatment effects (Angrist and Krueger 2001).

where, y_{ij} is weight-for-height or school enrollment for child i in family j, S_{ij} is a child sex dummy that takes a value of 1 for girls, θ_j represents a measure of the bargaining power of the mother in family j, d_j represents the female to total child ratio preferred by the mother in family j, r_j represents the actual female to total child ratio in family j, X_{ij} is a vector of observed child characteristics such as age and birth order, Z_j is a vector of observed household characteristics like household size and wealth status[32], μ_j represents unobserved family characteristics and ε_{ij} the error term that contains unobserved child characteristics and measurement error. We assume that μ_j and ε_{ij} are both normally distributed with zero means and constant variances σ_μ^2 and σ_ε^2, respectively.

The signs and magnitudes of δ_3 vs. δ_4 and δ_5 vs. δ_6 help us infer whether and how the deviation of the actual proportion of girls from the proportion desired by the mother along with her bargaining power influences health and education for boys and girls. δ_3 and δ_4 represent the effect of the gender ratio gap on boys and girls, respectively and if the compensating hypothesis proposed in this essay is true δ_3 should have a positive sign and δ_4 should have a negative sign; i.e., the bigger the excess of actual proportion of girls over the desired proportion, the less will the mother invest in the quality of each girl and the more will she invest in the quality of each boy to compensate for the deficit in the quantity of boys. δ_5 and δ_6 represent the effect of the interaction between gender ratio gap and the mother's power on investment in the human capital of boys and girls, respectively. If the compensating hypothesis is true, therefore, δ_5 has to be positive and δ_6 has to be negative but bigger in size relative to δ_3 and δ_4, respectively. On the other hand, if Becker's hypothesis that parents prefer equal distribution of quality across children is

[32] Controlling for family wealth status is important because fertility preferences and mother's bargaining power could differ by wealth class and its omission may bias our estimates for the key coefficients.

correct, all these coefficients should be close to zero since gender preferences, satisfied or not, should not affect the distribution of human capital.

As previously mentioned, there is a plausible reason why the signs of these coefficients in the schooling equation could even turn out to be the opposite of what is implied by the compensating hypothesis. For example, if the reason for mother's preference for more girls than boys is in order to help her with family activities at home, a girl born among boys may have less chance to be sent to school than a girl born among some boys and some girls. On the other hand, if the mother's preference for girls is a reflection of her wish to have daughters that are successful in their future life, then a girl born among boys is likely to have a better chance to attend a school than one born among some boys and some girls. Therefore, the empirical estimates of the coefficients will help us shed some light on the underlying cause of the mother's gender ratio preferences.

Estimation Methodology

There are two key issues that need be addressed in estimating the child health and schooling equations specified in the form of 2.6. First, our key variable of interest, mother's gender ratio gap, could be endogenous. While the actual gender composition of children (r_j) can be largely assumed to be exogenous (because of lack of technology for selective abortion in less developed countries like Ethiopia), the mother's gender ratio preferences (d_j) could arguably be correlated with the unobserved family characteristics

(μ_j). For example, the mother's preferences for gender mix[33] of children (d_j) could be influenced by the unobserved family characteristics in (μ_j) that may also influence her beliefs about the value of health and schooling for a girl vs. a boy. Therefore, valid instruments are needed to address the potential bias in the estimates that could arise because of the endogeneity. Second, the interaction of this potentially endogenous variable with other covariates calls for a modified application of the instrumental variables method.

We use two sources of identification for the mother's gender ratio gap. The first source is the gender ratio of the mother's own siblings that may have influenced her perceptions about and hence preferences for the gender mix of children but are not directly correlated with the health and schooling outcomes of her children. Conceptually, the gender ratio of the mother's siblings could influence her tastes for gender mix of her own children in at least two different ways. One possibility is that the mother may have liked the gender mix of her own siblings and may just wish to replicate it. It is also possible that the mother may not have liked the gender mix of her own siblings and may wish to have a different gender mix for her own children. The two may tend to offset each other and the gender mix of the mother's siblings may not be a very strong instrument for the mother's gender ratio gap. Therefore, we use the birth of living same-sex twins to the mother as an additional source of identification for the mother's gender ratio gap.

[33] In the demographic and health surveys that are used for analysis in this essay the question about the desired gender mix of children was asked retroactively by asking the mother how many boys, girls or children of any sex would she like to have if she could go back to the time she did not have any children.

The birth of same-sex twins can be treated as a random shock that substantially alters the gap between the desired and actual gender ratios just as the birth of twins is often treated as an exogenous shock that substantially changes the family size in unexpected way. The birth of twins has been used as instrument to identify the effect of family size on human capital and other outcomes of family members (e.g., Rosenzweig and Wolpin 1980; Black, Devereux and Salvanes 2007).The rationale for using the birth of twins as instrument for family size is that it can substantially deviate the actual family size from the desired family size (Black, Devereux and Salvanes 2007). Analogous arguments can be made for the use of the birth of same sex twins as instrument for the gender ratio gap in the health and enrollment equations specified in the previous section. At a given point in time parents decide to have the next child expecting that it can be a boy or a girl with equal probability. When they end up getting two girls or two boys instead, it is a shock that may substantially cause the actual gender ratio to deviate from the desired gender ratio in unexpected way. Therefore, the presence of female pairs of twins and male pairs of twins is used as additional instruments for the gender ratio gap in the health and enrollment equations.

To implement the instrumental variables estimation with an interactive endogenous regressor, we follow the approach formulated in Wooldridge (1997, 2003). The procedure involves generating instruments for the interactive terms and provides consistent estimators provided that the model satisfies the following conditions.

(1) The treatment variable must be continuous. This holds in our model since the potentially endogenous variable, gender ratio gap (r_j-d_j), is a continuous variable.

(2) The conditional expectation of the outcome variable, *y*, must be linear in the treatment variable. This also holds since y in equation 2.6 is specified to be linear in all of its covariates.

(3) The instruments must fulfill the standard exclusion restrictions in equation 2.6. That is, the instruments influence child health and schooling only through their effect on the gender ratio gap. The instruments we have selected for the gender ratio gap appear to fulfill the exclusion restrictions since there is no apparent way through which they can directly influence the outcome variables as previously argued.

Given these conditions, the Wooldridge's procedure essentially requires estimating a linear reduced form equation for the endogenous regressor and using the resulting fitted value and its interactions with the corresponding covariates in the model as new instruments. We estimate the child health and schooling models with this approach using cross-sectional data from siblings and twins that belong to married parents with at least two living children. We limit our sample to the currently married women with at least 2 children since these are the families where we can observe differential bargaining power between husband and wife and differences in outcomes for children, if any. The models can be estimated by OLS to obtain unbiased estimates correcting the standard errors for family level clustering. However, the paired structure of the data from twins, and to some extent siblings, provides additional information that can be exploited to obtain more accurate estimates (Carlin et al. 2005; Conley, Strully, and Bennet 2006). To see that, we can rewrite the error components in equation 2.6 as,

$$\varphi_{ij} = \alpha_j + \varepsilon_{ij} \tag{2.7}$$

The new composite error term, φ_{ij}, has zero mean and variance, $\sigma_\alpha^2 + \sigma_\varepsilon^2$. Apparently, φ_{ij} is not independently distributed within twin-pairs or within siblings since they share the same family level heterogeneity, μ_j. The within-twin or within-siblings correlation is $\sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma_\varepsilon^2)$ and the twins/siblings' random-effects GLS procedure that we use to estimate the models exploits this correlation to generate the required weights. The correlation between the errors is expected to be stronger for the twin-pairs since at least some[34] of the twins may share the hereditary components of the unobserved endowments like ability that are part of ε_{ij}. In addition, the twins' data allow us to control for the potential confounding effects of the over-time changes in the unobserved family characteristics and preferences since twins, unlike the ordinary siblings, face such changes together.

Data and Descriptive Statistics

The main empirical analysis in this essay is based on pooled cross-sectional data from two rounds of the Ethiopian demographic and health surveys (DHS)[35] conducted in the years 2000 and 2005 by the Central Statistical Agency of Ethiopia. Although the principal reason for pooling the two data sets is in order to obtain a reasonably large sample of twins, working with the pooled data should not be unreasonable for the purpose at hand since the key research question here does not closely relate to policy or other variables that could have changed between the two survey rounds. In any case, we try to generically control for the possible effects of the unobserved factors that might

[34] Both monozygotic and dizygotic twins are part of our sample but they are not identified in our data.
[35] The data were provided by the U.S. Agency for International Development which is the Financial Sponsor of all the Demographic and Health Surveys.

have changed between the two rounds by including a dummy for the survey round in our econometric models. As a way of checking the validity of our results in other contexts we also try to replicate our main results using data from the demographic and health survey of India conducted in 2005/06 with a bigger sample size.

The Ethiopian demographic and health survey of the year 2000 (DHS2000) was nationally representative survey of 14,072 households proportionately distributed across the 11 regions of the country. DHS2005 was a similar survey that covered 13,721 households. Both surveys consisted of three components: household questionnaire, women's questionnaire and men's questionnaire. The household questionnaire collected data on basic demographics and education for all individual members of the household, relationships among household members, and main assets owned by the household. The women's questionnaire was administered to all women of the reproductive age (15-49) in the sample households and asked reasonably detailed questions on fertility levels and preferences, child health and nutrition, as well as indicators of women's empowerment. The men's questionnaire addressed similar questions to a smaller number of men (2,607 men vs. 15,367 women in 2000 and 6,033 men vs. 14,070 women in 2005) but did not include questions on fertility history, child nutrition and health.

Most of the key variables of interest are available in both surveys. Data on educational attainment and attendance status are available for household members including the children. Child anthropometric data (weight and height) are available for under-5 children and the weight-for-height z-scores (waz) were calculated on the basis of the WHO standards. Information on the mother's preferences for the number and gender mix of children were collected by asking the mother how many children she would have liked to have "if she could go back to the time she did not have any children and could choose exactly the number of children to have in her whole life" and how many of those she would like to be boys, girls or any sex. This information is used to calculate the mother's gender ratio preference as the ratio of total number of girls preferred to total number of children preferred. The actual gender ratio is calculated by dividing the total number of living girls to the mother by the total number of her children. The gender ratio gap is the difference between the preferred proportion of girls and the actual proportion of girls at the time of the survey assuming no gap (a zero gap) when the mother does not care about the gender mix of her children.

To construct a measure(s) of women's bargaining power through factor analysis, we focus on the indicators that are available in both rounds. These include differences in schooling or literacy between husband and wife, age gap, religion, media exposure (frequency of listening to radio), presence of multiple wives (polygyny), and multiple self-reported indicators of independence and self-confidence of the mother in her relationship with her husband. All the indicators are scaled in an increasing order of importance for women's power so as to make the interpretation of the resulting factor(s) in our models easier.

In addition to the key variables of interest (gender ratio gap, mother's bargaining power and gender of the child), we include controls for the child's age, household size and birthorder as well as the households' relative wealth profile in our models. Controlling for the child's birthorder is important because both the child's physical health and the parents' choice to send him/her to school or not could depend on how many children they have before or after him. Controlling for family wealth status is also important because fertility preferences and mother's bargaining power could differ by wealth class and its omission may lead to biases in our estimates. The household's wealth quintiles (that are already available in the data) were constructed from the wealth indices that summarize the household's ownership of key assets like land and livestock. In the models for the child's physical health we also control for the mother's body mass index in order to account for some of the genetic variations in child's weight-for-height. The definitions of all the variables used in our econometric models including the instruments are presented in table 7 below.

Table 7. Descriptions of Variables Used in the Models

Dependent Variables	
Child is attending school	Dummy=1 if child is attending school
Weight-for-Height Z-scores	Z-scores for height standardized weight of the child
Covariates	
Mother's Desired Female Ratio	Desired number of girls/Desired Number of children
Actual Female Ratio	Actual number of girls/actual number of children
Gender Ratio Gap	Actual Female Ratio-Mother's Desired Female Ratio
Mother's Bargaining Power	A composite measure of mother's bargaining power from factor analysis
Gender Ratio Gap*Female	Gender Ratio Gap*Female
Gender Ratio Gap*Bargaining Power	Gender Ratio Gap*Bargaining Power
Gender Ratio Gap*Bargaining Power*Female	Gender Ratio Gap*Bargaining Power*Female
Child is female	Dummy=1 if child is female
Age of Child	Age of child in years
Household Size	Household size
Child's Birth-Order	Child's birth-order
First Wealth Quintile	Dummy=1 if household falls in the first wealth quintile
Second Wealth Quintile	Dummy=1 if household falls in the second wealth quintile
Fourth Wealth Quintile	Dummy=1 if household falls in the fourth wealth quintile
Fifth Wealth Quintile	Dummy=1 if household falls in the fifth wealth quintile
Mother's Body Mass Index	Mother's Body Mass Index

Survey Round is 2005	Dummy=1 if observation comes from the 2005 round
Instruments for Gender Ratio Gap	
Gender Ratio of Mother's Siblings	Number of mother's sisters/Number of mother's siblings
Female Twin-Pair in the Family	Dummy=1 if the mother has a living female pair of twins
Male Twin-Pair in the Family	Dummy=1 if the mother has a living male pair of twins
Mixed-sex Twin-Pair in the Family (control group)	Dummy=1 if the mother has a living mixed sex pair of twins
Indicators of mother's bargaining power used in factor analysis	
Independence in Seeking External Relations	Dummy=1 if mother thinks wife beating by husband is unjustified for not telling him where she is going
Independence in Child-care Decisions	Dummy=1 if mother thinks wife beating by husband is unjustified for failing in caring for children
Self-confidence in Talking to Husband	Dummy=1 if mother thinks wife beating by husband is unjustified for arguing with him
Independence in Sexual Decisions	Dummy=1 if mother thinks wife beating by husband is unjustified for refusing to have sex with him
Independence in Domestic Activities	Dummy=1 if mother thinks wife beating by husband is unjustified for burning food
Husband has just 1 wife	Dummy=1 if the mother is the only wife of her husband
Mother Listens to Radio	=0 if not at all, =1 if less than once a week, =2 if at least once a week, =3 if almost every day
Mother is More Educated than Husband	Dummy=1 if mother is more educated than husband
Age-gap	Mother's age-Husband's age
Mother is Non-Muslim	Dummy=1 if mother is non-Muslim

Source: The Demographic and Health Surveys.

The focus of analysis in this essay is on the children belonging to the currently married women with at least 2 children since these are the families where we can observe differential bargaining power between husband and wife and differences in outcomes for children, if any. Our sample for schooling consists of the biological children of such parents who were 6-18 years old during the survey rounds while our models for child’s physical health are estimated using data from the under-5 year olds belonging to such families. The summary statistics computed from the combined data for the variables used in our models are presented in table 8 below. Table D1 in appendix D presents the descriptive statistics for the two rounds separately.

We have a total of 23,819 children with complete data in our sample for schooling out of which 252 or 126 pairs are twins. On the other hand, our sample for the analysis of the child's physical health consists of 9,504 children with complete data out of which 148 or 74 pairs are twins. The total number of twin pairs belonging to married couples with at least two children is 268 (84 female pairs, 76 male pairs and 108 mixed sex pairs) including those with missing age and/or anthropometric data and those older than 18. While our regression results for twins are based on those who have complete data for the key variables, the gender of twin-pairs as an instrument for the gender ratio gap is based on all the living twin-pairs. In our sample for schooling, about 0.9%, 1% and 1.4% of the children belonged to families with living female-twin pairs, male-twin pairs and mixed-sex-twin pairs, respectively. On the other hand about 0.9%, 0.7% and 1.1% of the children in our sample for the child-health models belonged to families with living female-twin pairs, male-twin pairs and mixed-sex-twin pairs, respectively. While this is a small sample of twins in absolute terms, it is generally comparable to the twins' samples used for the purpose of instrumenting for family size in some other studies. For example, the percentage of twins in the sample used by Rosenzweig and Wolpin (1980) was 1.5%.

Table 8. Summary Statistics for the Variables Used in the Models

	School Attendance (N=23819)		Weight-for-Height (N=9504)	
Variables	Mean	Std. Dev.	Mean	Std. Dev.
Child is attending school	0.369	0.483		
Weight-for-Height Z-scores			-1.719	1.294
Percentage of wasted children			0.436	0.496
Mother's Desired Female Ratio	0.470	0.150	0.471	0.162
Actual Female Ratio	0.485	0.229	0.491	0.267
Gender Ratio Gap	0.015	0.214	0.020	0.247
Mother's Bargaining Power (de-meaned)	0.000	0.886	0.000	0.881
Child is female	0.478	0.500	0.497	0.500
Age of Child	10.204	3.213	2.129	1.409
Household Size	7.239	2.005	6.424	1.998
Child's Birth-Order	3.831	2.344	4.637	2.528
First Wealth Quintile	0.201	0.401	0.177	0.381
Second Wealth Quintile	0.170	0.376	0.172	0.378
Fourth Wealth Quintile	0.189	0.391	0.200	0.400
Fifth Wealth Quintile	0.237	0.425	0.235	0.424
Mother's Body Mass Index			20.239	2.540
Gender Ratio of Mother's Siblings	0.489	0.252	0.476	0.242
Female Twin-Pair in the Family	0.009	0.093	0.009	0.096
Male Twin-Pair in the Family	0.010	0.098	0.007	0.083
Mixed Twin-Pair in the Family	0.014	0.116	0.011	0.105
Independence in Seeking External Relations	0.368	0.482	0.384	0.486
Independence in Child-care Decisions	0.335	0.472	0.334	0.472
Self-confidence in Talking to Husband	0.388	0.487	0.364	0.481
Independence in Sexual Decisions	0.467	0.499	0.451	0.498
Independence in Domestic Activities	0.391	0.488	0.355	0.478
Husband has just 1 wife	0.843	0.364	0.871	0.335
Mother Listens to Radio	0.499	0.895	0.460	0.850
Mother is More Educated than Husband	0.058	0.233	0.066	0.249
Mother's age-Husband's age	-9.036	6.784	-8.555	6.654
Mother is Non-Muslim	0.592	0.491	0.609	0.488
Survey Round is 2005	0.475	0.499	0.324	0.468

Source: Combined Data from the Demographic and Health Surveys of Ethiopia Conducted in the Years 2000 and 2005.

The evidence in table 8 shows that only 37% of the children in the combined data from the two rounds were attending school at the time of the surveys. The attendance rates between the two survey rounds were not very far apart as shown in table D1 in appendix D (35% in 2000 vs.39% in 2005). The average height-for-age z-score of about -1.72 also shows a high prevalence of child wasting in Ethiopia. According to WHO standards[36] a child with less than -2 weight-for-height z-score is considered to be wasted. According to this criterion, therefore, about 44% of the children were wasted. Although we observe some improvement in the percentage of children wasted in the 2005 round (46% in 2000 vs. 39% in 2005) it is important to note that the number of children with anthropometric data in 2005 is less than half of that in 2000 and some of it could be statistical discrepancy as demonstrated by larger standard deviation in 2005. The relationships between the variations in child schooling and health outcomes and the proposed covariates are presented in the next section.

Estimation Results

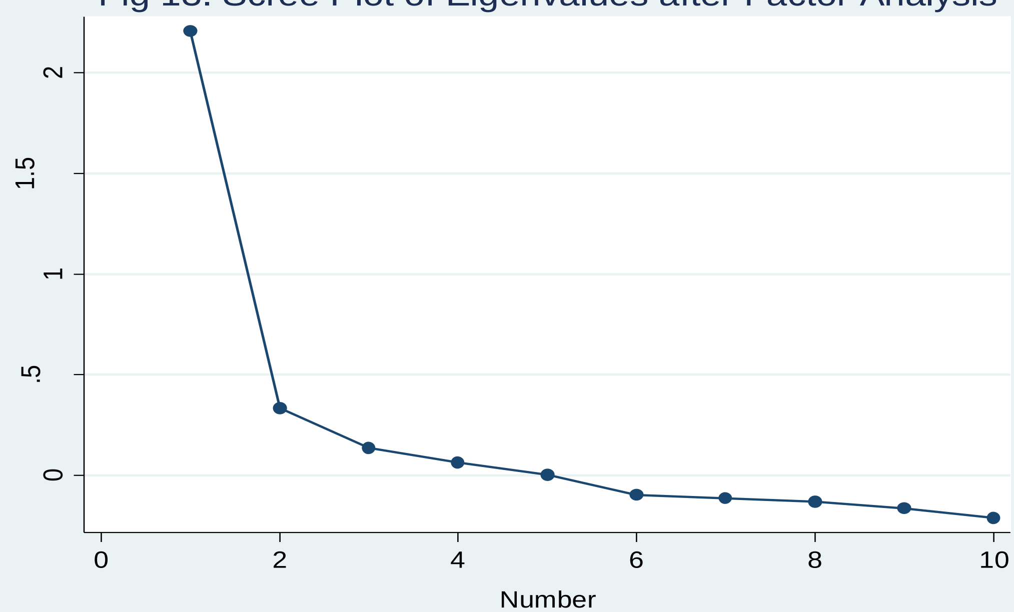
Factor Analysis

As previously stated we use factor analysis to construct an index (indices) of mother's bargaining power that focus on the common variation of selected indicators. The factor loadings and the corresponding eigenvalues of the correlation matrix between the 10 selected indicators are presented in table D2 and table D3 in appendix D, respectively. The scree plot of the eigenvalues for the 10 factors is presented as Fig 13

[36] The WHO child growth standards are available at http://www.who.int/childgrowth/standards/weight_for_height/en/index.html, accessed April, 2010.

below. The magnitudes of the eigenvalues reported in table D3 indicate that none of the factors except the first capture substantial common variation in the selected indicators. Only the first factor has an eigenvalue well in excess of 1 after which it drastically drops to 0.34 for the second factor and levels off as demonstrated in Fig 13 below. Therefore, both the Kaiser-Guttman rule and Cattell's scree test imply that only the first factor contains important information about the common variation in the selected indicators. We thus take factor1 as an indicative index of women's decision making power in the family and use it as a covariate in our child health and schooling models.

Fig 13. Scree Plot of Eigenvalues after Factor Analysis



All the 10 indicators of the mother's power are positively correlated in factor1 as demonstrated by their positive factor loadings reported in table D2. This is in fact, by design since all the indicators were scaled in an increasing order of importance for women's power. Given the positive correlation between all the components of factor1, therefore, we can interpret its increasing values as signifying increasing mother's power and the vice versa. This is so because those women who have larger values of the constituent indicators will also have larger values of the index, other things remaining the same.

Although factor1 accounts only for about 22.1%[37] of the total underlying variation in the indicators, it is capturing almost all of the common variation in the indicators since the other factors account for negligible fraction of the common variation in the variables. Whether the amount of variation in the original indicators we have captured in factor1 is adequate for our purpose or not mainly depends on how well the index performs in our models. Most of the results reported in the next section show that the index contains important information about women's power.

It is also important to note that factor1 mostly captures the variation in the 5 self-reported indicators of mother's independence and self-confidence in the relationship with her husband. The other indicators have relatively small but non-negligible common variation with the self-reported indicators of empowerment and hence are kept in the model. While it is possible to argue that the self-reported indicators are the direct measures of mother's empowerment whose influence on child health and schooling come through the latter, some of the other indicators like radio access and education gap could also have direct effects on child health and schooling. For example, a mother who listens to radio frequently may be better informed not only about her rights in the family and community (empowerment) but also about the benefits of child schooling and health. However, the part of the variation in radio access that is captured by our index is the part

[37] This is obtained by dividing the sum of the squared factor loadings by the number of variables which is the same as dividing the eigenvalue by the number of variables.

that is positively correlated with the self-reported direct indicators of empowerment. In other words, the role of access to radio will be captured in factor1 only when women who frequently listen to radio are also more independent and self-confident in their relationship with husbands. Therefore, it is less likely that our index of mother's bargaining power will be reflecting the direct effects of access to radio and the other indicators on child health and schooling.

Results for Child Schooling

In this section we present the estimation results for the models of schooling. As previously stated our main estimation procedure involves two stages. We first regress our potentially endogenous variable, the gender ratio gap, on the instruments and other covariates in our models using OLS. The fitted value from this first stage regression and its interactions with the corresponding covariates are then used as instruments for the gender ratio gap and its interactions. In the second stage, therefore, we are estimating an exactly identified equation since the number of generated instruments is exactly equal to the number of endogenous covariates including the interactive terms. One benefit of this approach is that we don't need to correct the standard errors in the second stage since we are using generated instruments instead of generated regressors as Wooldridge (2003) states.

The first stage results for our models of child schooling are reported in table D4 in appendix D. The gender ratio of the mother's siblings as an instrument is positively correlated with the mother's gender ratio gap both in the siblings' and twins' data and it is statistically significant in the case of the siblings' data. The birth of female twins has significant negative effect on the mother's gender ratio gap while the birth of male twins has the opposite effect. The instruments are jointly significant in both the siblings' and the twins' models as demonstrated by the F-test for their joint significance.

We estimate the models both with OLS and random-effects GLS including their instrumental variables counterparts. In most of the cases, the OLS and GLS estimates including their instrumental variables counterparts for our key variables of interest are similar in sign and close in magnitude but the GLS estimates generally have smaller standard errors than their OLS counterparts. However, the instrumental variables estimates for our key variables are often larger than their simple OLS and GLS counterparts perhaps implying that endogeneity of the gender ratio gap was indeed an issue that needed to be addressed. Therefore, our subsequent discussion mainly focuses on the instrumental variables GLS estimates.

The first set of results we report as baseline estimates exclude the mother's bargaining power and its interaction with the gender ratio gap and just focus on the latter. These results are reported in tables D6 and D7 in appendix D for the siblings' and twins' data, respectively. The results from the siblings' data indicate that the mother's gender ratio gap has highly significant positive effect on the probability of school attendance for boys and marginally significant negative effect on the probability of school attendance for girls. This implies that the bigger the excess of actual proportion of girls over the mother's desired proportion, the smaller a girl's chance of attending school and the bigger a boy's chance of attending school will be. According to these results a girl born to a mother who desires to have 3 boys and 3 girls will have a bigger chance of attending school when she is born with 5 brothers than 2 sisters and 3 brothers. These results are consistent with our compensating hypothesis and are replicated using similar[38] but larger data from India as reported in table E4 in appendix E. However, we do not find similar results from the twins' data perhaps due to the relatively small sample of twins available in the data sets. Therefore, the support for our compensating hypothesis in our baseline estimates for schooling appears to be mixed at best.

Table 9. Linear Probability Models for School Attendance-Results from Siblings' Data

VARIABLES	(1) OLS	(2) GLS	(3) 2SLS	(4) IV-GLS
Child is female	-0.056***	-0.062***	-0.158***	-0.099***
	(0.006)	(0.006)	(0.030)	(0.020)
Gender Ratio Gap	0.038*	0.058***	1.055***	0.992***
	(0.023)	(0.022)	(0.286)	(0.303)
Mother's Bargaining Power	0.057***	0.057***	0.029**	0.026*
	(0.004)	(0.004)	(0.014)	(0.015)
Gender Ratio Gap*Female	-0.010	-0.040	-0.546	-0.745**
	(0.030)	(0.028)	(0.372)	(0.375)
Gender Ratio Gap*Bargaining Power	-0.085***	-0.083***	-0.523**	-0.545**
	(0.024)	(0.023)	(0.231)	(0.256)
Gender Ratio Gap*Bargaining Power*Female	0.067**	0.066**	0.822**	0.872**
	(0.033)	(0.031)	(0.374)	(0.360)
Age of Child	0.043***	0.043***	0.043***	0.042***
	(0.001)	(0.001)	(0.001)	(0.001)
Household Size	-0.001	-0.001	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.003)
Child's Birth-order	-0.007***	-0.006***	-0.009***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
First Wealth Quintile	-0.146***	-0.142***	-0.147***	-0.142***
	(0.011)	(0.011)	(0.010)	(0.012)
Second Wealth Quintile	-0.038***	-0.038***	-0.045***	-0.042***
	(0.012)	(0.011)	(0.010)	(0.013)
Fourth Wealth Quintile	0.041***	0.040***	0.043***	0.043***
	(0.011)	(0.011)	(0.010)	(0.012)
Fifth Wealth Quintile	0.206***	0.205***	0.206***	0.207***
	(0.011)	(0.011)	(0.010)	(0.012)
Survey Round is 2005	0.145***	0.145***	0.151***	0.149***
	(0.007)	(0.007)	(0.007)	(0.008)
Constant	-0.098***	-0.099***	-0.019	-0.030
	(0.018)	(0.017)	(0.036)	(0.037)

[38] The only difference is that the set of instruments in the Indian case does not include the gender ratio of the mother's siblings since such data were not gathered in the demographic and health survey of India in 2005/06.

Observations	23819	23819	23819	23819
Number of Mother_id		9911		9911

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The sample consists of children from age 6 to 18 in the combined data from the two rounds of DHS-Ethiopia. Only those children with married mother with at least two children are included.

The main set of results we report for child schooling include the index of mother's bargaining power and its interaction with the gender ratio gap as additional covariates to the baseline equations. These results are presented in tables 9 and 10 for the siblings' and twins' data, respectively. It is important to note that for the effects of the interactions between the mother's bargaining power and the gender ratio gap to be identified there must be differences in their gender preferences. Although we cannot verify this for the entire sample since data on fertility preferences were not collected for all the men, the small subsample of men for whom the data are available (2168 in 2000 and 4752 in 2005) appear to have substantially different gender preferences than women (see table D10 in appendix D). In the DHS2000, the preferred number of boys by the husband and wife differs in 77% of the cases while the preferred number of girls differs in 73% of the cases. In DHS2005, the corresponding figures are 76% and 72% for DHS2000 and DHS2005, respectively.

The results from the siblings' data demonstrate that the index of the mother's bargaining power we constructed through the factor analysis has significant positive effect on the probability of child schooling which is generally consistent with the findings in the existing literature. However, it is important to note that the magnitude and sign of the coefficient of the mother's bargaining power that shows up along the third row in table 9 shows its effect that comes independent of its interaction with the mother's gender ratio gap and hence may not tell the full story about the role of the mother's power. It is also important to note that the absolute magnitudes of the coefficients of the interactive terms are not very informative since they reflect the combined effect of multiple variables. Our focus is thus on the signs and statistical significances of the coefficients for boys and girls.

Table 10. Linear Probability Models for School Attendance-Results from Twins' Data

	(1)	(2)	(3)	(4)
VARIABLES	OLS	GLS	2SLS	IV-GLS
Child is female	0.061	-0.019	-0.104	-0.104
	(0.061)	(0.051)	(0.096)	(0.096)
Gender Ratio Gap	-0.200	-0.087	0.408	0.408
	(0.199)	(0.198)	(0.637)	(0.637)
Mother's Bargaining Power	0.034	0.041	-0.012	-0.012
	(0.048)	(0.050)	(0.057)	(0.057)
Gender Ratio Gap*Female	0.216	0.089	0.899	0.899
	(0.244)	(0.193)	(0.861)	(0.861)
Gender Ratio Gap*Bargaining Power	-0.569*	-0.421	-1.287**	-1.287***
	(0.335)	(0.299)	(0.497)	(0.497)
Gender Ratio Gap*Bargaining Power*Female	0.754*	0.620*	1.515**	1.515**
	(0.399)	(0.355)	(0.754)	(0.754)
Age of Child	0.049***	0.044***	0.051***	0.051***
	(0.011)	(0.012)	(0.010)	(0.010)
Household Size	-0.015	-0.015	-0.010	-0.010
	(0.018)	(0.018)	(0.016)	(0.016)
Child's Birth-Order	-0.004	-0.022	0.003	0.003
	(0.014)	(0.015)	(0.013)	(0.013)
First Wealth Quintile	-0.144	-0.157	-0.121	-0.121
	(0.123)	(0.135)	(0.109)	(0.109)
Second Wealth Quintile	-0.011	-0.047	-0.080	-0.080
	(0.112)	(0.120)	(0.126)	(0.126)
Fourth Wealth Quintile	0.071	0.072	0.056	0.056
	(0.138)	(0.151)	(0.127)	(0.127)
Fifth Wealth Quintile	0.257**	0.237*	0.261**	0.261***
	(0.108)	(0.121)	(0.101)	(0.101)
Survey Round is 2005	0.159**	0.136*	0.135*	0.135*
	(0.078)	(0.082)	(0.073)	(0.073)
Constant	-0.182	0.024	-0.246	-0.246
	(0.180)	(0.208)	(0.214)	(0.214)
Observations	252	252	252	252
Number of Mother_id		121		121

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: The sample consists of living twin-pairs from age 6 to 18 in the combined data from the two rounds of DHS-Ethiopia. Only those children with married mother with at least two children are included.

In the results for both siblings and twins, the signs and statistical significances of the coefficients of the gender ratio gap for boys and girls are similar to those of the baseline line estimates although the statistical significance of our preferred IV-GLS estimate of the coefficient for girls has improved in the results for the siblings' data. On the other hand, the sign of the coefficient for the interaction between the gender ratio gap and the mother's bargaining power is negative and significant for boys and positive and significant for girls in both the siblings' and twins' results. These results[39] are replicated using Indian data as reported in tables E8 and E9 in appendix E and imply that a girl born to an influential mother has a bigger chance of attending school when more than desired proportion of girls is born.

These results apparently contradict our expectation that the mother's bargaining power will reinforce the effect of the gender ratio gap in the direction implied by the compensating hypothesis but make intuitive sense in the context of Basu's (2006) hypothesis. According to this hypothesis, too much mother's power could be unfavorable to a child's attendance of schooling in the contexts where the child is needed to help out with family activities. While Basu's hypothesis is not gender specific as formulated, if it is true, it is likely that a girl born to a powerful mother (who would like to keep some of the children for family labor activities) will be more demanded at home when she is the only girl among 5 boys than she is one among 3 boys and 3 girls. This is so because boys

[39] We obtain similar results when we use a dummy taking a value of 1 for the positive values of the index for the mother's bargaining power and 0 otherwise in the place of the index itself.

and girls often have different roles in family labor activities and having siblings with similar gender could be an advantage for a child in terms of attending school since he/she can share the burden of those gender specialized activities with them. Therefore, when fewer than the mother's desired proportion of girls or boys are born, their chances of attending school could be diminished most likely because mothers use their power to retain them to help in household activities. This in itself implies that the mother's desire to have more girls/boys is not necessarily because she wants to have daughters/sons that are successful in their future life but rather because she wants them to help in specific family activities. The bottom line is thus, the mother's power is important in influencing the distribution of schooling among her children but could be unfavorable to boys' or girls' chances of schooling depending on their gender mix relative to her preferences. Although we do not want to stretch the policy implications of these results too far before they are confirmed in other contexts with different data sets and estimation techniques, they seem to indicate that women empowerment programs will have to be supplemented with other policies that mitigate their needs for child labor (like income support programs) in order to enhance child schooling.

Results for Child's Physical Health

We estimate the equations for child's physical health, proxied by weight-for-height z-scores, using similar procedures followed in our models for schooling. The first stage results for our models of child's health are reported in table D5 in appendix D. As in the case of our models for schooling the instruments are jointly significant in the first stage equations of the models for child health as demonstrated by the large values of the

F-stat in table D5. Once again we first estimate the models for child's health by ignoring the role of the mother's bargaining power. The results for these models are reported as baseline in tables D8 and D9 in appendix D for the siblings' and twins' data, respectively.

The baseline results from the siblings' data presented in table D8 indicate that the mother's gender ratio gap has a positive effect on the weight-for-height of boys and negative effect on the weight-for-height of girls. The signs of the corresponding estimates from the twins' data presented in table D9 are consistent with the results from the siblings' data for girls but our preferred IV-GLS estimate for boys turns out to be negative in the results for twins. In the baseline results from Indian data reported in tables E6 and E7 in appendix E, the IV-GLS estimate of the coefficient of the mother's gender ratio gap is positive for boys and negative for girls both in the siblings' and twins data'. Therefore, the balance of these results appears to indicate that there is some support to the compensating hypothesis that the birth of the less than desired proportion of children of either gender (particularly girls) may be augmented by favorable nutrition to the children of that gender. However, most of these coefficients are statistically insignificant and the results have to be cautiously interpreted.

Table 11. Models for Child's Weight-for-Height Z-Scores -Results from Siblings' Data

VARIABLES	(1) OLS	(2) GLS	(3) 2SLS	(4) IV-GLS
Child is female	-0.010	-0.006	0.301	0.154
	(0.028)	(0.027)	(0.389)	(0.282)
Gender Ratio Gap	0.021	0.045	1.544	1.428
	(0.086)	(0.084)	(1.236)	(1.232)
Mother's Bargaining Power	0.041**	0.044**	-0.068	-0.006
	(0.017)	(0.017)	(0.120)	(0.104)
Gender Ratio Gap*Female	-0.064	-0.100	-5.023*	-3.780
	(0.120)	(0.118)	(2.882)	(2.506)
Gender Ratio Gap*Bargaining Power	-0.184*	-0.161*	-1.240	-0.646
	(0.096)	(0.093)	(1.320)	(1.163)
Gender Ratio Gap*Bargaining				

Power*Female	0.219*	0.184	2.182	1.100
	(0.132)	(0.129)	(2.259)	(1.936)
Age of Child	-0.180***	-0.173***	-0.192***	-0.181***
	(0.009)	(0.009)	(0.011)	(0.011)
Household Size	0.037***	0.037***	0.018	0.023*
	(0.009)	(0.009)	(0.014)	(0.014)
Child's Birth-Order	-0.049***	-0.049***	-0.084***	-0.075***
	(0.007)	(0.007)	(0.021)	(0.018)
First Wealth Quintile	0.021	0.029	0.006	0.016
	(0.045)	(0.045)	(0.047)	(0.049)
Second Wealth Quintile	-0.041	-0.033	-0.015	-0.014
	(0.043)	(0.043)	(0.048)	(0.049)
Fourth Wealth Quintile	0.055	0.057	0.063	0.063
	(0.041)	(0.041)	(0.044)	(0.046)
Fifth Wealth Quintile	0.249***	0.258***	0.259***	0.274***
	(0.041)	(0.040)	(0.044)	(0.045)
Mother's Body Mass Index	0.091***	0.091***	0.095***	0.094***
	(0.006)	(0.006)	(0.008)	(0.008)
Survey Round is 2005	0.280***	0.277***	0.248***	0.258***
	(0.030)	(0.030)	(0.039)	(0.038)
Constant	-3.339***	-3.373***	-2.982***	-3.082***
	(0.125)	(0.125)	(0.219)	(0.219)
Observations	9504	9504	9504	9504
Number of Mother_id		6425		6425

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The sample consists of children under age 5 in the combined data from the two rounds of DHS-Ethiopia. Only those children with married mother with at least two children are included.

Our main models for the child’s physical health expand the baseline line models by including the index for the mother’s bargaining power and its interactions with the gender ratio gap as additional covariates. The results for these models are presented as tables 11 and 12 for the siblings’ and twins’ data, respectively. The signs of the coefficient of the gender ratio gap for boys and girls are consistent with the compensating hypothesis in these results as well, but we observe indications of the mother’s bargaining power reinforcing the compensating hypothesis only in the twins’ data. In the results for twins presented in table 12 below the effect of the interaction between the mother’s bargaining power and the gender ratio gap is positive for boys and negative for girls while it takes the opposite signs in the results from the siblings' data. However, none of these coefficients is statistically significant in the IV-GLS equations.

Table 12. Models for Child's Weight-for-Height Z-Scores -Results from Twins' Data

VARIABLES	(1) OLS	(2) GLS	(3) 2SLS	(4) IV-GLS
Child is female	-0.088	0.101	0.339	0.382
	(0.221)	(0.166)	(0.401)	(0.323)
Gender Ratio Gap	1.790**	0.444	-0.078	0.145
	(0.864)	(0.695)	(2.643)	(3.391)
Mother's Bargaining Power	0.154	0.156	0.317	0.353
	(0.181)	(0.172)	(0.246)	(0.305)
Gender Ratio Gap*Female	-2.845**	-0.864	-5.166	-6.390
	(1.308)	(1.115)	(5.668)	(8.018)
Gender Ratio Gap*Bargaining Power	1.437*	1.387*	4.324**	4.193
	(0.821)	(0.772)	(1.870)	(2.560)
Gender Ratio Gap*Bargaining Power*Female	-0.788	-0.220	-5.892	-7.568
	(1.491)	(1.517)	(5.666)	(7.329)
Age of Child	0.209**	0.200**	0.275**	0.285*
	(0.092)	(0.090)	(0.118)	(0.151)
Household Size	0.011	0.049	-0.004	-0.017
	(0.101)	(0.108)	(0.076)	(0.121)
Child's Birth-Order	-0.092*	-0.116**	-0.160*	-0.144**
	(0.052)	(0.054)	(0.089)	(0.072)
First Wealth Quintile	-0.394	-0.532	-0.829	-0.617
	(0.570)	(0.570)	(0.666)	(1.057)
Second Wealth Quintile	-0.673	-0.687	-0.866*	-0.709
	(0.447)	(0.484)	(0.449)	(0.736)
Fourth Wealth Quintile	1.149**	0.923	0.876	1.085
	(0.565)	(0.576)	(0.679)	(1.091)
Fifth Wealth Quintile	1.023**	0.830*	0.737	0.951
	(0.439)	(0.450)	(0.528)	(0.898)
Mother's Body Mass Index	0.060	0.083	0.062	0.036
	(0.053)	(0.055)	(0.090)	(0.139)
Survey Round is 2005	-0.013	0.051	-0.170	-0.225
	(0.376)	(0.381)	(0.421)	(0.587)
Constant	-3.554**	-4.213***	-3.023	-2.633
	(1.405)	(1.397)	(1.923)	(2.921)
Observations	148	148	148	148
Number of Mother_id		73		73

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The sample consists of living twin-pairs under age 5 in the combined data from the two rounds of DHS-Ethiopia. Only those children with married mother with at least two children are included.

In the results from Indian data reported in tables E10 and E11 in appendix E the signs of the coefficient of the interaction between gender ratio gap and the index for mother's power are consistent with the compensating hypothesis in the IV-GLS equations but the estimates are statistically insignificant for both boys and girls. Based on these results, therefore, there is little evidence in support of the claim that the mother's power may be used to reinforce the effect of the gender ratio gap in the direction implied by the compensating hypothesis when it comes to child's nutrition health. Perhaps the mother doesn't want to discriminate between her children in terms of something as basic as food no matter how the realized gender mix deviates from what she would have liked.

Conclusion

In this essay we try to empirically investigate how the gap between mother's preferred and actual gender mix of children influences the distribution of schooling and nutrition health across boys and girls. Specifically, we try to examine whether the quantity deficit in the children of the preferred gender is compensated by their favorable treatment in terms of investment in schooling and nutrition (that we call compensating hypothesis) and to what extent the mother uses her bargaining power in the family to influence this process. We mainly use data from siblings and twins in two rounds of the demographic and health surveys of Ethiopia, although we replicate the results using data from a similar but larger data from India.

An index of mother's power in the family is constructed from multiple indicators using factor analysis and the potential endogeneity of the gap between the mother's preferred and actual gender mix of children is instrumented for by the gender mix of the mother's own siblings as well as the birth of female and male pairs of twins to the mother. Our models for child schooling and nutritional health (measured in terms of child's weight-for-height) are estimated using both OLS and random-effects GLS. Our preferred results are those based on the random effects GLS method that exploits the paired structure of the siblings' and twins' data to produce more accurate estimates than OLS.

We found no conclusive evidence that the mother tries to compensate the deficit in the quantity of children of her preferred gender by more favorably investing in their schooling and nutrition although there are qualitative indications in some of our results that this could be the case. While our expectation was that higher mother's power will reinforce the effect of mother's gender ratio gap on the distribution of child schooling and nutritional health in the direction implied by the compensating hypothesis, we rather find it working in the opposite direction in the case of child schooling and having no substantive role in the case of child nutritional health. In other words, our results for schooling imply that a girl born to an influential mother has a bigger chance of attending school when more than desired proportion of girls is born than not.

Our results for schooling, though unanticipated, make some intuitive sense in the context of Basu's (2006) hypothesis about the role of mother's power in child labor and schooling. According to this hypothesis, too much mother's power could be unfavorable to a child's attendance of schooling in the contexts where the child is needed to help out with family activities. While Basu’s hypothesis is not gender specific as formulated, it is likely that a girl born to a powerful mother (who would like to keep some of the children for family labor activities) will be more demanded at home when she is the only girl among 5 boys than she is one among 3 boys and 3 girls. This is so because boys and girls often have different roles in family labor activities and having siblings with similar gender could be an advantage for a child in terms of attending school since he/she can share the burden of those gender specialized activities with them.

Therefore, when fewer than the mother's desired proportion of girls or boys are born, their chances of attending school could be diminished because the mother uses her power to retain them to help in household activities. This in itself implies that the mother’s desire to have more girls/boys is not necessarily because she wants to have daughters/sons that are successful in their future life but rather because she wants them to help in specific family activities. The bottom line is thus, the mother’s power is important in influencing the distribution of schooling among her children but could be unfavorable to boys’ or girls’ chances of schooling depending on their gender mix relative to her preferences. Although we do not want to stretch the policy implications of these results too far before they are confirmed in other contexts with different data sets and estimation techniques, they seem to indicate that women empowerment programs will have to be supplemented with other policies that mitigate the need for child labor (e.g., income support programs) in order to enhance child schooling.

ESSAY III: TEENAGE SOCIALIZING BEHAVIOR AND SCHOOLING OUTCOMES FOR AMERICAN YOUTH

Introduction

Recent empirical studies have shown that there is some decline in the ‘actual’ high school graduation rates and some ‘slow-down’ in the college enrollment rates in the U.S. While the official estimates based on ‘status completion’ show high school graduation rates as high as 88% and somewhat rising (Heckman and Lafontaine 2010), various studies have used alternative methods and data sources to come up with much lower and slightly declining graduation rates over the last three decades. For example, Miano and Haney (2004) report national high school graduation rates ranging between 66% and 80% for the period from 1973 to 2001, slightly declining particularly after early 1990s. Heckman and LaFontaine (2010) focused on earning formal high school diploma (i.e., not including those with a general educational development certificate, GED) as an appropriate measure of high school graduation and compared alternative data sets to conclude that the graduation rates based on the latter are much lower than the officially reported rates and have in fact declined over the recent years[40]. There is also substantial variation in the high school graduation rates by race and gender, men generally falling behind and the outcomes being worse for blacks and Latinos. They associate the observed recent slow-down in college enrollment and the rising female-male college attendance ratio to this declining high school graduation rate that disproportionately affects men.

[40] The distinction between high school graduation with the standard diploma and just GED is important because there is now a substantial literature, cited in Heckman and Lafontaine (2010), indicating that the GED provides far lower returns than the standard high school diploma.

A number of empirical studies in the past tried to investigate the socio-economic factors that influence schooling outcomes in general and high school graduation rates in particular. The mainstream economics literature in this area largely focused on factors unrelated to the choices that the child him/herself makes. These include family income and financial constraints, family background, and demographic and school characteristics as possible determinants of child education. Other studies have examined the influence of child endowments and environmental factors that are largely beyond the control of the family and the child. However, the differences in school outcomes across individuals could not be fully explained by factors beyond the control the individual himself/herself. As a result recent studies in labor economics have expanded the literature on the determinants of child schooling outcomes by examining the effects of some choices that children themselves make such as involvement in crime, drug use, alcohol consumption, teenage pregnancy and child birth. These behaviors have been subject to empirical and theoretical research in other social science disciplines like sociology and psychology for a long time, but it is relatively recently that rigorous methodologies of economics have been applied to identify causal relationships between these variables and schooling outcomes.

What appears to be largely missing from the economics literature on the relationship between the choices made by the child and schooling outcomes is the direct effect of participation in sexual and dating activities. Despite the fact that Becker (1980) developed a broad theoretical model that could be applied to almost any rational choice that a person makes, the empirical researchers in economics appear to so far be reluctant to expand the literature on the determinants of child schooling in this direction. I could find no empirical study that examines the influence of the frequency and intensity of teenage dating on schooling outcomes. There are a few studies, particularly in sociology and psychology that tried to examine the effect of early involvement in sex on schooling (see the next section). However, these studies are based on simple correlations and fail to recognize the possible endogeneity of the teenagers' decisions to engage in sexual activities in the models for educational outcomes. Hence, their estimates might not represent causal effects. There are two recent publications (Sabia 2007a, 2007b) that try to address the issue of endogeneity of teenage sex in school outcome equations, but the author uses a narrowly defined sex variable (loss of virginity) and the results are inconclusive . The purpose of this essay is, therefore, to examine the effect of teenage involvement in dating and sexual activities on schooling outcomes, using a broader set of measures that capture the intensity of the teenager's participation in these activities. In addition, an attempt is be made to use a new set of instruments and control variables from a different data set (NLSY97) to address the problem of endogeneity.

Literature

The bulk of economics literature on the determinants of school attainment in the past emphasized the effect of family income and financial constraints on schooling outcomes. For example, Cameron and Heckman (1998) examined the determinants of grade by grade schooling attainment for cohorts of American males born between 1908 and 1964 and found a strong correlation between family income and transitions from one level of schooling to the next starting from elementary school to graduate school. Similar findings were reported by a number of other studies like Mayer (1997), Levy and Duncan

(2000) and Cameron and Heckman (2001). This may seem to indicate that expanding access to credit would enhance educational attainment by the needy. However, a number of other studies found no evidence that borrowing constraints play a major role in explaining the educational attainment gap between children from high and low income families. For instance, in a model where schooling choices and returns to schooling are jointly determined, Cameron and Taber (2004) analyzed the impact of borrowing constraints on the years of schooling attained by the sample of American youth in the 1979 NLSY cohort and found no evidence that borrowing constraints play a role. Similar findings were reported by Keane and Wolpin (2001) and Carneiro and Cameron (2002). The latter argue that access to credit may play only a minor and short-term role while the educational gap between the children from the high and low income family backgrounds is mostly explained in terms of the long-term family effects. Clearly, much more than family income and financial constraints must be involved in explaining differences in educational attainments.

Another area of focus for empirical research on educational attainment of children has been on the effect of parents' education. Parental education to some extent is related to the effect of income and wealth but even beyond income there could be some 'hereditary' and 'environmental' effects (Keane and Wolpin 2001) given that children born from better educated parents may put higher value on educational achievement and learn from their parents' success. While the conventional wisdom (with some empirical support) is that mother's education is much more important than father's education (Haveman and Wolfe 1995), there are some recent studies that cast some doubt on this perception. Using data from twins, Behrman and Rosenzweig (2002) have found that mother's education has little effect on the child's education if the unobserved abilities (like the hereditary elements) are held constant. Similarly, Plug (2004) used evidence from adopted children and finds "no treatment effect for the mother's schooling, conditional on the husband's schooling" (p.366) and suggested that the effect of parent's schooling on children's education could largely be hereditary reinforced by "positive assortative mating". These findings indicate that the usual argument about the value or quality of more educated mother's time being more productive in producing child's human capital does not survive the control for endowment or genetic effects. However, these studies were based on data from samples drawn from atypical populations (Farré, Klein, and Vella 2009) and further studies may be needed to confirm their findings.

Other family background variables, whose effects on educational attainments have been widely investigated by various studies summarized in Haveman and Wolfe (1995), include parents' labor force status and occupation, wealth, race, age, marital status, number of siblings, birth order, ethnicity, language, and urban/rural residence. The effects of some government assistance programs (like AFDC or TANF), neighborhood characteristics, school types and amenities have also been tested (see Haveman and Wolfe 1995). Other studies have focused on differences in endowments (e.g., ability) to explain differences in child school achievements. While some studies (e.g., Behrman, Rosenzweig and Taubman 1994) find endowment differences could matter in schooling attainments, others (e.g., Ashenfelter and Krueger 1994) found "no evidence that unobserved ability is positively related to the schooling level completed" (p.1157). Therefore, factors unrelated to the choices that the child makes explain some but not all of the variation in educational achievements among children.

As such, it should not come as big surprise that the economics literature on the determinants of school attainment has lately expanded to incorporate variables that are directly related to the choices that the child himself/herself makes like involvement in crime, drug use, alcohol consumption and smoking. These may affect educational achievements not only because of their effect on cognitive and physical health through addiction, but also because of their potential effect on social status. For example, Register, Williams and Grimes (2001) try to analyze the effect of drug use on the number of years of schooling completed and find strong negative correlation. Chatterji (2006) finds similar results after accounting for endogeneity of drug use. DiSimone and Wolaver (2005) find alcohol consumption to have important negative effect on educational attainment for risk averse students after accounting for unobserved heterogeneity. There are some sociological studies like Jenkins (1995) that try to estimate the effect of delinquent behavior on school outcomes. The reverse effect of education on such outcomes has also been investigated by some studies (e.g., Kenkel 1991). Of course, any serious investigation of the effects of such variables has to address the apparent simultaneity between these and the schooling outcomes but many studies, including some recent ones like Register, Williams and Grimes (2001), fail to do so.

One issue related to the child's choice whose effect on educational attainment has long attracted attention is teenage pregnancy and child bearing. There exists a voluminous empirical literature on this issue including but not limited to Bronars and Grogger (1994), Geronimus and Korenman, 1992), Hoffman, Foster, and Furstenberg (1993), Ahn (1994), Klepinger ,Lundberg, and Plotnick (1999), Hofferth, Lori, and Frank (2001) and Hotz, McElroy, and Sanders (2005). The overwhelming evidence is that teenage pregnancy and childbearing have a strong negative effect on educational attainment although the effects are smaller once unobserved heterogeneity is accounted for. However, there could be other social, psychological and physiological reasons why teenage engagement in dating and sexual activities could affect schooling outcomes even when it does not result in pregnancy and childbearing. For example involvement in such activities may lead to reduced focus on long-term objectives, changing priorities in time use, and changing attitudes towards academic achievement (Billy et al., 1988). These reasons may be particularly relevant when there are multiple sex and dating partners and an increasing intensity of engagement in such activities.

There are a few studies, particularly in sociology and psychology, that tried to examine the effect of early involvement in sex on schooling outcomes (e.g., Mott and Marsiglo 1985; Billy et al. 1988; Schvaneveldt et al. 2001; Rector and Johnson 2005). The findings are mixed but these studies are based on simple correlations failing to recognize the possible endogeneity of the teenagers' decisions to engage in sexual activities in the models for academic achievement and hence their estimates may not represent causal effects. For example, it might be the case that those who opt to spend a lot of time in dating and sex are those who are less capable and hence have limited prospects of doing well at school.

In the economics literature, two recent studies by Sabia (2007a, 2007b) recognize the potential endogeneity of teenage sexual activities in the models for school achievement and tried to correct for the potential bias. Using data from National Longitudinal Study of Adolescent Health, the author attempted to control for possible unobserved heterogeneity using fixed effects and instrumental variables estimation techniques. He concludes that losing virginity by teenagers has negligible or no effect on school attachment and achievement in the form of GPA once controlling for the endogeneity. However, his results might not be conclusive given the narrow definition of involvement in sexual activity and school achievement he adopted.

The detailed data on teenage sexual activity available in the NLSY97 data set makes it possible to test the effect of not only the loss of virginity as in Sabia (2007a, 2007b) but also the effect of the intensity of sexual activity in the form of number of times the teenager had sex over a period, number of sex partners and the age at which the teenager had sex for the first time. The data set also contains information on the age of first-dating with a boy friend or girl-friend, frequency of dating and number of people dated. Therefore, this essay tries to estimate the effect of each of these on the schooling outcome variables controlling for teenage pregnancy and child birth, cohabitation and early marriage, whose effects in turn may be confounded by the failure to account for the effects of sex and dating.

Theoretical Background

In Becker's (1991) single-person family utility maximization model, a child's (teenager's) decision problem involving schooling, dating, sex and other sources of satisfaction can be modeled treating schooling as one of the z-goods for the child. Since the focus of this essay is on the role of the choices made by the child himself/herself, the parental preferences and resources are assumed to be part of the environmental variables for the child. Ignoring sex for the time being, suppose the teenager generates utility directly from schooling(S), dating (D), and (Z)-all other sources of welfare including leisure, sleeping, etc.

$$U=U(S, D, Z) \tag{3.1}$$

The child is both producer and consumer of S, D and Z and the production of each of these requires material inputs purchased in the market (x) as well as time inputs (t) from the child. For example material inputs of schooling will include books and material inputs of dating will include drinks. In other words, the production functions for S, D, and Z will look like,

$$S=f_s(x_s, t_s; A) \tag{3.2}$$

$$D=f_d(x_d, t_d; A) \tag{3.3}$$

$$Z=f_z(x_z, t_z; A) \tag{3.4}$$

where A represents factors like the child's ability, motivation, psychological health, plus environmental variables such as parental preferences and resources that the child cannot directly control. Substituting (3.2), (3.3) and (3.4) into (3.1) the utility function of the child becomes,

$$U = U[f_s(x_s, t_s; A), f_d(x_d, t_d; A), f_z(x_z, t_z; A)] \tag{3.5}$$

Suppose p_i^x represents the market price of the x inputs where $i=s,d,z$ and w represents the opportunity cost of time for the teenager; i.e., w is the wage rate he/she could earn if he/she were to work (or whatever the valuation placed on activities foregone). Also suppose the teenager spends a total time of t_w working. Then the total time constraint is,

$$t_s+t_d+t_z+t_w=t \tag{3.6}$$

and the full income budget constraint for the child will be,

$$p_s^x x_s + p_d^x x_d + p_z^x x_z + w(t_s + t_d + t_z) = w(t - t_s - t_d - t_z) + v = F \tag{3.7}$$

Here, v is the non-labor income of the child including the direct transfers from the parents that the child can decide upon how to spend. Maximizing (5) subject to (7) with respect to x_i and t_i gives us the optimal amount of purchased and time inputs into each of the commodities as a function of the predetermined variables as,

$$x_i^* = x_i(p_s^x, p_d^x, p_z^x, w, F, A), \quad \text{where } i=s, d, z \tag{3.8}$$

$$t_i^* = t_i(p_s^x, p_d^x, p_z^x, w, F, A), \quad \text{where } i=s, d, z \tag{3.9}$$

Substituting (3.8) and (3.9) into (3.2), (3.3) and (3.4) gives us *S*, *D* and *Z* as a function of the material input prices, the full income (*F*) and the endowment and environmental variables (*A*). However, we can use the input prices to calculate the costs of producing the commodities that directly enter the child's utility function (*S*,*D* and *Z*) that represent their shadow prices(π) as,

$$\pi_s = p_s^x \frac{x_s^*}{S} + w\frac{t_s^*}{S} \tag{3.10}$$

$$\pi_d = p_d^x \frac{x_d^*}{D} + w\frac{t_d^*}{D} \tag{3.11}$$

$$\pi_z = p_z^x \frac{x_z^*}{Z} + w\frac{t_z^*}{Z} \tag{3.12}$$

Then, the full income budget constraint can be written in terms of the shadow commodity prices as,

$$\pi_s S + \pi_d D + \pi_z Z = F \tag{3.13}$$

Taking the other commodities (*Z*) as the base category, we can obtain the demand functions[41] for *S* and *D* as functions of shadow prices, the full income and the environmental and endowment variables as,

$$S^* = S(\pi_s, \pi_z, F, A) \tag{3.14}$$

$$D^* = S(\pi_d, \pi_z, F, A) \tag{3.15}$$

A demand function similar to (3.15) can also be derived for involvement in sexual activities by a child following the same procedures. Clearly, the equations for schooling and dating are highly interdependent not only because of the way the shadow prices are defined but also because of the presence of common endowment and environmental variables in both equations. Since time allocations have to add up to a fixed time endowment as per equation (3.6) and since dating and schooling may compete for the

---

[41] The detailed characterizations of the demand functions in the *z*-goods context are available in Becker (1991).

same material inputs (x), their shadow prices are interdependent. The nature of the interdependence may depend on whether dating and schooling are substitutes or complements to each other. For example, if dating involves studying together and sharing school materials with your dating-mate, then dating and schooling could be considered complements to each other and increased allocation of time and resources for one could increase production of the other as well. However, if the purpose of dating is something unrelated to schooling including romance and hanging out together, then dating could lead to less time, attention and motivation for studying and may result in diminished schooling outcome. In addition, dating may affect one's mood, degree of happiness or motivation (based, for example, on one's partner's attitude toward schooling) that can either improve (complement) or worsen (substitute) school outcomes. Therefore, the effect of dating on schooling outcomes is theoretically ambiguous. Depending on whether the complementarity or substitution effect is stronger, we may observe negative, positive or no empirical correlation between indicators of dating and schooling outcomes. Similar arguments can be made for involvement in sexual activities. The effects of the substitutes and complements for dating and sex such as cohabitation, marriage, pregnancy and having children come into play through their shadow prices (π_z).

Methodology

The focus of analysis in this essay is on how high school completion is affected by involvement in dating and sex as a teenager. To define an empirical model for high school completion, suppose s_i* represents the indirect utility for child i obtained by substituting the optimal amounts of schooling and other activities into the child's utility function described in the previous section. The indirect utility is unobserved to the researcher but whether the child graduates from high school or not is observed and assumed to depend on whether or not this maximized utility takes at least a minimum threshold value. The indirect utility depends on the shadow prices of schooling and other activities that represent opportunity costs, anticipated returns to schooling, the full income as well as child endowment and environmental variables. Suppose s_i* is linearly related to these factors such that,

$$s_i^* = X_i \beta + e_i \tag{3.16}$$

where, X_i represents a vector of observable variables like shadow prices of schooling, shadow prices of dating or sex and their substitutes and complements, child's full income including transfers from parents, and other parental resources. Unobservable endowments and environmental variables like motivation and ability fall into the error term e_i that is assumed to be normally distributed with mean 0 and variance 1. Now, let s_i represent high school completion status of individual *i* that takes a value of 1 if the maximized utility s_i^* is positive and zero otherwise i.e.,

$$s_i = \begin{cases} 0 \text{ if } \mathrm{s_i}^* \leq 0 \\ 1 \text{ if } \mathrm{s_i}^* > 0 \end{cases} \tag{3.17}$$

The distribution of s_i conditional on X_i can then be modeled as probit or can be approximated by a linear probability model. If observations on all the components of X_i including the shadow prices of dating or sex and their substitutes and complements were available and exogenous, we could regress the dichotomous outcome for high school completion (s_i) on these covariates and try to infer the effects of dating or sex variables on the basis of the estimates for their component variables. While theoretically the shadow prices could be calculated from the observed input prices as described in the previous section, in practice this is impossible because observations on the inputs and their prices are unavailable. Alternatively, therefore, we treat the observed dating and sex variables as direct covariates in the equations for schooling, but recognize that they are endogenous.

One of the possible reasons for the endogeneity is that both school outcomes and the decision to engage in dating and sexual activities may be influenced by common unobserved individual and family characteristics. For example, it might be the case that those who choose to spend a lot of time in dating and sex are those who are less capable and hence have limited prospects of doing well at school. If this is the case the simple OLS estimates will tend to overestimate the causal effects of the teenage dating and sex on schooling. Another source of bias in the OLS estimates could be the under-reporting of engagement in sexual/dating activities by the teenagers. The under-reporting may lead to understatement of the effects of teenage dating and sex on schooling outcomes. The various sources of bias may thus tend to offset each other.

The specific variables that are used to represent involvement in sexual activity are age at which the teenager had sex for the first time, number of sex partners over ages 15 to 17[42] and number of times the teenager had sex over the same period. Similarly,

---

[42] The selection of this age range is essentially dictated by the fact that the oldest members of the sample were age 16 during the first round and hence age 15, 16 and 17 are the teenage years for which information on these variables is available for everybody in the sample (data for age 15 is available for the oldest respondents because questions relating to these variables were asked for the last year).

engagement in dating activities are represented by age at first date with a girlfriend or boy friend, number of dating partners and frequency of dating over ages 15 to 17. The effect of each of these variables on high school completion is cumulative in nature since a diminished performance in earlier grades, say because of too many sex/dating partners, will make it less likely that the student will complete high school.

Each of the indicators of involvement in sex/dating are treated as endogenous covariates (one at a time) in the schooling equations. Controls for teenager's choices or their outcomes that could be substitutes or complements to sex and/or dating like cohabitation, marriage, and teenage child birth are included in the models since sex and dating variables could simply be capturing the effects of these variables. In addition, an indicator for above or below mean family income is included as a proxy for the availability of parental resources. We also include parental education, race and residence at age 17 as additional controls.

The fact that the sex or dating variables are endogenous in the models for schooling means that identification of their causal effect requires the existence of variables that influence the taste for sex or dating but do not directly affect schooling. The data from the 1997 cohort of the National Longitudinal Survey of Youth (NLSY97) that are used for analysis in this essay contain information on the frequency of church visits by parents, and percent of peers who go to church regularly that may influence the teenager's inclination towards engaging in sexual activities[43] but do not seem to have apparent direct effect on schooling outcomes. Therefore, indicators of parental and peers'

[43] See Evans, Oates and Schwab (1992) for peer group effects and L'Engle, Christine, and Brownet al. (2006) and Brewster (1998) for the effect of religion on teenage sex.

religiosity are used as instruments for teenage sex and dating variables, in an attempt to estimate their causal effects on schooling outcomes.

Although all the endogenous dating and sex variables used here take discrete (integer) values and hence the continuity and normality assumptions required for consistent estimates through application of instrumental variables probit[44] to such cases may not strictly hold, linear approximations of such models provide consistent estimates of the average treatment effects (Angrist and Krueger 2001). Therefore, our main results are based on linear probability models for high school completion and college enrollment although for our key models we also report the marginal effects from probit estimates in the appendix. For our main models we test the instruments for exogeneity using Sargan's (1958) and Basmann's (1960) chi-squared tests for over-identification.

It is worth noting that most of the control variables included in the models like cohabitation, marital status, and child bearing might also be endogenous. As a result the estimated coefficients of such variables may not necessarily represent their causal effect on the school outcome variables. Our principal interest, however, is not in estimating and interpreting the causal effects of these variables since that has been done elsewhere, but rather to control for the potential omitted variable bias in the estimated coefficients of dating and sex variables that may arise as a result of their correlation with these control variables. As such, the properly estimated coefficients of the variables of interest are expected to be valid even when some of the other regressors in the models are

[44] In the instrumental variables Probit model, the errors in the first and second stage equations are assumed to have a bivariate normal distribution. This means that the endogenous regressor has to have features of a normal random variable. Discrete variables do not strictly fulfill this requirement due to lack of continuity (for details see Wooldridge, 2002:472-478).

endogenous. The control variables are sequentially introduced into the models so that we can see how much of the observed relationship between the specific sex or dating experience indicator and the school outcome variable was due to other observable variables that are correlated with sex or dating.

Finally, we estimate models for college enrollment to see if involvement in dating and sexual activities as a teenager has any lasting effect beyond high school completion. The covariates and estimation procedures for the college enrollment models are similar to those in the models for high school completion, but the college enrollment outcome is made conditional on high school completion. For both high school completion and college enrollment, we also estimate and report separate models for boys and girls to see if the effects of teenage dating and sex vary by gender. The models for girls include teenage pregnancy as an additional control variable since teenage pregnancy could have lasting psychological or physiological effects even if it doesn't lead to child bearing. The definition and summary statistics for all the variables used in the models are presented in the next section.

Data and Descriptive Statistics

Analysis in this essay is based on data from the 1997 cohort of the National Longitudinal Survey of Youth (NLSY97). The original sample for this cohort consisted of 8984 young men and women aged 12 to 16 on December 31, 1996. The data from the first 11 rounds of the survey are publicly available and contain detailed information on the individual and family characteristics as well as processes and outcomes for the youth.

The key outcome variables we are interested in are whether the youth have completed high-school at or before age 19 and subsequently enrolled in college at or before age 20. The weighted summary statistics for these and other variables used in the models are presented in table 13 below. The summary statistics disaggregated by gender and race are presented in tables F1 and F2 in appendix F, respectively. Our estimation sample consists of 6026 youth with complete data for all the variables of interest out of which 3054 are girls and the balance are boys.

As briefly stated in the introductory section, there is no consensus as to how to measure high school graduation rate. Status completion rates that include the recipients of the GED and certificates of attendance often show a much higher graduation rate than the rates that count only those who have received formal high school diploma. This is true in our sample as well 88.7% of the youth having completed high school while only 81.8% have formal high school diplomas. The high school completion rate obtained for this sample of youth is very close to the estimated completion rate of 88% issued by the National Center for Educational Statistics (NCES). For our econometric analysis we focus on the high school completion that includes GED recipients instead of only graduates with diploma since the GED is "generally accepted as the equivalent of a high school diploma for college admissions" (Heckman and LaFontaine 2010).

The summary statistics for boys and girls presented in table F1 in appendix F show that both the high school completion rate and high school graduation rate (with a diploma) are higher for women than men, the gap being slightly bigger in the latter case with 83.8% of the girls having a high school diploma compared to 79.7% of boys. There is also substantial racial disparity in both the high school completion rate and graduation-with-diploma rate as presented in table F2. While the outcomes are essentially identical for blacks and Latinos, the completion rate is about 9 percentage points higher for the whites and the graduation rate is about 10 percentage points higher. The gender and racial disparities observed in these data are consistent with the general pattern in the U.S. (and other developed countries) that girls on average do better in terms of high school and undergraduate educational outcomes and minorities on average do worse.

Table 13. Description and Weighted Summary Statistics for the Variables Used in the Econometric Models

Variable	Description	Mean	Std. Dev.
Schooling Outcome			
High School Completion	Dummy=1 if completed high school at/before age 19	0.887	0.317
High School Diploma	Dummy=1 if graduated with high school diploma	0.818	0.386
College Enrollment	Dummy=1 if enrolled in college at or before age 20	0.542	0.498
Enrollment | HS Completion	Dummy=1 if enrolled in college at or before age 20 given high school completion	0.609	0.488
Teenage Sex Indicators			
Had Sex Under 15	Dummy=1 if had sex under 15	0.181	0.385
Had Sex Under 18	Dummy=1 if had sex under 18	0.513	0.500
Sex Partners	Total Number of sex Partners from age 15 to 17	4.420	9.528
Age at 1st sex	Age at first sex	16.09	2.101
Total Sex	Total Number of times had sex from age 15 to 17	134.5	298.9
Teenage dating indicators			
Had Date Under 15	Dummy=1 if had date under 15	0.599	0.490
Had Date Under 18	Dummy=1if had date under 18	0.936	0.245
Dating Partners	Total Number of dating Partners from age 15 to 17	13.64	21.00
Age at 1st Date	Age at first date	14.04	1.954
Total Dates	Total Number of times had dates from age 15 to 17	25.93	87.94
Controls for Family Background and demographics			
Male	Dummy=1 if male	0.496	0.500
Black	Dummy=1 if race is black	0.156	0.363
Hispanic	Dummy=1 if race is Hispanic	0.059	0.236
White (excluded)	Dummy=1 if race is white	0.723	0.448
Dad college educated	Dummy=1 if dad is college educated	0.529	0.499
Mom college educated	Dummy=1 if mom is college educated	0.511	0.500
Above mean income	Dummy=1 if family earns above mean income	0.464	0.499
Rural at age 17	Dummy=1 if rural resident at age 17	0.281	0.449
Control for teenage marriage and cohabitation			

Married under 18	Dummy=1 if married under age 18	0.159	0.366
Cohabited under 18	Dummy=1 if cohabited under age 18	0.290	0.454
Controls for teenage pregnancy and child bearing			
Pregnant under 18 (girls)	Dummy=1 if ever pregnant under age 18	0.216	0.411
Had kids under 18	Dummy=1 if had kids under age 18	0.133	0.339
Instruments			
Parental church visit1	Dummy=1 if parent visited church once or less per month in 1997	0.505	0.500
Parental church visit2	Dummy=1 if parent visited church twice per month in 1997	0.121	0.326
Parental church visit3	Dummy=1 if parent visited church once a week in 1997	0.264	0.441
Parental church visit4	Dummy=1 if parent visited church several times a week in 1997	0.108	0.310
Parental church visit5	Dummy=1 if parent visited church every day in 1997	0.003	0.054
Peer church visit	Dummy=1 if more than 50% of peers visited church in 1997	0.260	0.439

Source: Various rounds of NLSY97.
Note: N=6026, Number of Girls=3054. About 5.7 of the sample are from other races. In the regression equations these are included in the excluded category (whites).

In terms of college attendance, the summary statistics in table 13 show that about 54.2% of the youth were enrolled in college at or before age 20. The college enrollment rate for high school completers (including those with the GED) is about 7 percentage points higher than the total sample which is in line with Heckman and Lafontaine's (2010) argument that low high school graduation is one of the reasons for the observed slowdown in college attendance over much of the last decade. Once again girls in these data do much better in terms of college attendance with the conditional college enrollment rate of 65.7% compared to 55.8% for boys as shown in table F1. The enrollment rate for non-Hispanic whites is about 10 percentage points higher than that of blacks and about 7 percentage points higher than that of Hispanics. The overall enrollment rate of about 61% for the high school completers in our data is close to the enrollment rates in the CPS data for the late 1990's and the early years of the last decade as demonstrated in fig 14 below.

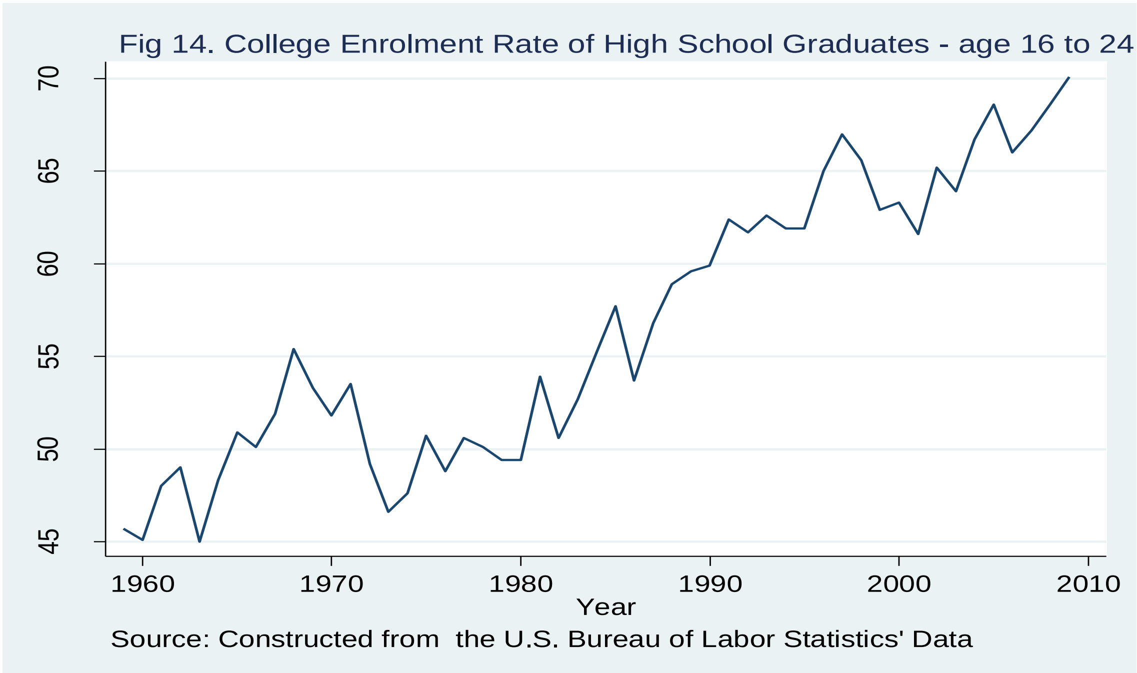


Table 13 also presents summary statistics for sexual and dating experience of the respondents as teenagers. The evidence shows that about 51.3% of the respondents have reported to have had sex under age 18 including 18.1% who have already begun sex under 15. A bigger percentage of boys appears to have started sex under 15 than girls although the proportion is roughly the same for boys and girls for the under 18 sexual experience. Boys appear to be more sexually active on average than girls in terms of number of sex partners but girls seem to have higher average frequency of sex over ages 15 to 17 than boys although the standard errors are quite large in both cases. The fact that boys appear to start sex earlier than girls is somewhat surprising given that girls mature earlier than boys and men generally date younger girls. It is possible that girls are more shy or secretive about revealing their early sexual experience or boys could be defining sex and dating more loosely than girls.

Among the racial groups, blacks on average seem to start sex earlier, 64% reporting to have had sex under 18 and 32% under 15 whereas the corresponding figures for Hispanics and whites are 56 (20) and 50 (15), respectively. Blacks have the largest number of sex partners on average but reported to have had sex the smallest number of times on average over ages 15 to 17 while whites had the opposite experience.

The teenage dating experience follows a similar general pattern to that of teenage sex experience in terms of gender disparity in the number of dating partners and starting age. According to the evidence presented in table F1, girls on average start dating later than boys and keep much smaller number of dating partners. About 66.6% of boys and 53.3% of girls reported to have started dating under age 15 while 93.7% of boys and 93.5% of girls already started dating under age 18 with an average of 11 partners for girls and 17 partners for boys over ages 15 to 17. Boys also have slightly higher average frequency of dating than girls. Unlike the case with teenage sex where a larger fraction of blacks start sex under 15, a relatively larger fraction of whites (followed by Hispanics) appear to start dating under 15 and keep it slightly more intense than blacks in terms of number of dating partners and frequency of dating.

It is important to note that the mean age at first sex/date reported in the tables is less informative given that it is top-coded at 18; i.e., everybody who reported to have had first sex/date at age 18 or later or didn’t yet have sex/date during the latest round are all recorded at 18. We chose to top-code age at first sex/date at 18 since our interest is to examine the effect of sex/dating as a teenager on schooling outcomes. In the regression equations, age at first sex/date is expected to capture the effect of waiting at least up to age 18 and still account for the effect of the variation in the sex/dating–start age as a teenager (unlike the under 18 sex/dating dummy that doesn't capture this variation).

The summary tables also contain descriptive statistics for the control variables including teenage marriage, cohabitation, pregnancy and child birth. We observe a much higher prevalence of teenage marriage and cohabitation among girls which is not surprising given their much quicker biological maturity than boys. There is also a much higher prevalence of teenage child birth among girls than boys with 20.1% of girls having a kid under age 18 compared to just 6.4% of boys. In terms of racial disparities, we observe a higher rate of teenage pregnancy and child birth among black girls than whites and Hispanics, and higher prevalence of teenage marriage or cohabitation among Latinos. The effects of these and the multiple indicators of teenage sex and dating on high school completion and college enrollment are discussed in the next section.

Estimation Results

As stated in the methodology section, the equations for high-school completion[45] are estimated treating the indicators of teenage sex and dating as endogenous regressors. Each of the multiple indicators of teenage involvement in sex and dating activities enters the regression equations separately since they mostly contain overlapping information and we do not have sufficient instruments to identify their effects all at once. The controls

[45] The qualitative aspects of the results remain the same when we use high school graduation with a formal diploma instead of the status completion as the dependent variables but the magnitudes of the estimates are mostly larger in the case of high school diploma.

for teenage pregnancy and child bearing as well as marriage and cohabitation are sequentially introduced into the equations so that we can see how much of the observed relationship between the specific sex or dating variable and the schooling outcome variable was due to other observable variables that are correlated with sex or dating.

Due to the discrete nature of our key variables of interest we prefer to focus on the linear probability models for high school completion that provide consistent estimates even when the first stage is mis-specified. Robust standard errors are used to correct for the inefficiency arising from heteroskedastic errors in the linear probability models. The marginal effects from probit models corresponding to our key linear probability estimates are also reported in appendix F. The instruments are tested for over-identification using Sargan and Basmann's chi-squared test and the first stage results for our main models are reported in appendix F.

While the summary statistics reported in the previous section were weighted using the sampling weights, the regression results reported and discussed in this section are unweighted. Using sample weights in the regressions is generally recommended when the interest is in obtaining the estimates representative of the entire population. Weighted estimators, however, tend to be more variable than the unweighted estimators as demonstrated for example in Korn and Graubard (1995). Hence, sample weights were not used in the regressions reported here as also suggested in the guidelines for the use of sampling weights in NLSY97[46]. As such, the estimates reported in this section may not

[46] See the guidelines at http://www.nlsinfo.org/nlsy97/nlsdocs/nlsy97/use97data/weights.html, accessed July, 2010.

strictly represent the characteristics of the entire U.S. population since variable probabilities of selection into the sample were used for various groups.

Teenage Sex and High Completion

The linear probability estimates of the effects of multiple indicators of teenage involvement in sexual activities are presented in table 14 below. The complete results for the models summarized in table 14 are reported in tables F6-F8 in appendix F and the first stage results for the 2SLS models are reported in tables F3 to F5 in the same appendix. The first stage results show that the instruments are both individually and jointly significant in the equations for all the three indicators of teenage sex. The results show that parental and peer religiosity is an important constraint to involvement in sexual activities by teenagers. Frequency of parental and peer church visits have negative effects on the number of sex partners and frequency of teenage sex but positive effect on age at first sex. In addition, the Sargan and Basmann chi-squared test for over-identification does not reject exogeneity of these instruments in the models for high school completion. Therefore, instruments capturing parental and peer religiosity appear to be valid for teenage sex variables in the models for high school completion.

The estimates for the effects of the number of teenage sex partners on high school completion are reported along the first row of table 14. While the coefficient of the number of teenage sex partners is negative in sign and highly significant in all of the OLS and 2SLS estimates, its magnitude is more than 15 times bigger in the case of 2SLS. Controlling for teenage marriage and cohabitation only slightly increases the magnitude of the 2SLS estimate while controlling for teenage child bearing slightly reduces its magnitude. Our 2SLS estimate from the model where all controls are included indicates that increasing ones sex partner between ages 15 to 17 by 1 reduces the probability of completing high school by 4.6% on average. The corresponding marginal effect at the mean from a probit model has similar sign and statistical significance but slightly smaller magnitude (3.7%) as presented in table F17 in appendix F. Therefore, maintaining large number of sex partners as a teenager appears to have non-negligible negative effect on the probability of graduating from high school even after controlling for teenage marriage and child birth.

Table 14. Teenage Sex and High School Completion by Age 19-
Estimates from Linear Probability Models

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.003***	-0.003***	-0.002***	-0.047***	-0.049***	-0.046***
	(0.001)	(0.001)	(0.001)	(0.010)	(0.014)	(0.013)
Age at 1st sex	0.028***	0.025***	0.022***	0.099***	0.101***	0.095***
	(0.002)	(0.002)	(0.002)	(0.014)	(0.017)	(0.017)
Total sex /100	-0.010***	-0.008***	-0.006***	-0.113***	-0.120***	-0.115***
	(0.002)	(0.002)	(0.002)	(0.020)	(0.026)	(0.027)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	6026	6026	6026	6026	6026	6026

Robust standard errors in parentheses
***Significant at 1%; ** significant at 5%; * significant at 10%
Notes: Each of the teenage sex indicators separately entered into the regression equations.

It appears that there is also some benefit to delaying initiation of sex as demonstrated by a highly significant positive coefficient for age at first sex as shown along the second row of table 14 above. As it is the case with the number of sex partners, the 2SLS estimate of the coefficient of age at first sex is much bigger than the OLS estimate and shows little change when controls for teenage marriage, cohabitation and child birth are included. After including all the control variables, our estimate shows that delaying initiation of sex by an additional year increases the probability of high school completion by 9.5%. The corresponding marginal effect at the mean from a probit model has similar positive sign and statistical significance but slightly bigger magnitude (11.4%) as shown in table F17 in appendix F.

Another indicator of teenage sex whose effects we analyzed is the number of times the teenager had sex from age 15 to 17. Since the frequencies are somewhat large for some of the youth and the absolute magnitudes of the effects of single sexual encounters are small, we rescaled the observed values of this variable dividing them by 100 for convenience in presenting the results. As shown along the third row of table 14 above, the frequency of teenage sex also has highly significant negative effect on high school completion with little changes in the coefficient when we control for teenage marriage and child bearing. Our preferred estimate obtained after controlling for these variables shows that having sex 100 times as teenager may reduce the probability of completing high school by 11.5% with almost identical (11.3%) marginal effect at the mean from a probit model.

In tables 15 and 16 below we present estimates for boys and girls[47] separately to see if there are gender differences in the effects of teenage involvement in sexual activities. In models for girls we include teenage pregnancy as an additional control

---

[47] An attempt was also made to see if results vary by racial groups by separately estimating the models for blacks, Hispanics and whites. For whites (n=3480) the results are both qualitatively and quantitatively similar to those we obtained for the total sample. For blacks (n=1578) and Hispanics (n=622), the signs of the estimated coefficients are largely similar to those we obtained for the total sample, but the magnitudes mostly differ and the estimates are mostly statistically insignificant except age at first sex for blacks. The same pattern is observed in the results for teenage dating by the racial groups.

variable. According to these results, the number of teenage sex partners has bigger effect on high school completion of girls than boys while age at first sex and frequency of teenage sex have bigger effect on the outcome for boys. Focusing on the model in which we control for all the relevant variables of interest, increasing the number of teenage sex partners by 1 reduces the probability of high school completion for girls by 5% compared to 3.5% for boys. On the other hand, delaying age at first sex by an additional year increases the probability of high school completion for girls by 6.2% compared to 10.3% for boys. Similarly, the effect of the frequency of teenage sex on the probability of high school completion is more than twice higher for boys than girls.

Table 15. Teenage Sex and High School Completion for Girls by Age 19- Estimates from Linear Probability Models

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.004**	-0.002	-0.000	-0.062***	-0.065***	-0.050**
	(0.002)	(0.001)	(0.001)	(0.015)	(0.020)	(0.021)
Age at 1st sex	0.030***	0.025***	0.013***	0.083***	0.080***	0.062***
	(0.004)	(0.004)	(0.004)	(0.015)	(0.019)	(0.022)
Total sex /100	-0.009***	-0.005**	-0.001	-0.092***	-0.089***	-0.071**
	(0.002)	(0.002)	(0.002)	(0.022)	(0.028)	(0.031)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	3054	3054	3054	3054	3054	3054

Robust standard errors in parentheses
***Significant at 1%; ** significant at 5%; * significant at 10%
Notes: Each of the teenage sex indicators separately entered into the regression equations.

It is also important to note that controlling for teenage pregnancy and child birth in the equations for girls substantially reduces the magnitudes of the coefficients of each of the teenage sex indicators as shown along the last column of table 15. On the other hand, controlling for teenage child birth for boys leaves the coefficients of the teenage sex indicators unchanged as shown in the last column of table 16. This implies that some of the effects of teenage sex for girls come through the resulting pregnancy and child-bearing while for boys the effects are essentially unrelated to teenage child birth.

Table 16. Teenage Sex and High School Completion for Boys by Age 19- Estimates from Linear Probability Models

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.003***	-0.002***	-0.002***	-0.034***	-0.035**	-0.035**
	(0.001)	(0.001)	(0.001)	(0.012)	(0.014)	(0.014)
Age at 1st sex	0.025***	0.023***	0.022***	0.102***	0.103***	0.103***
	(0.003)	(0.003)	(0.003)	(0.024)	(0.027)	(0.027)
Total sex /100	-0.012***	-0.010***	-0.009***	-0.133***	-0.149***	-0.150***
	(0.003)	(0.003)	(0.003)	(0.036)	(0.046)	(0.046)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	2972	2972	2972	2972	2972	2972

Robust standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the teenage sex indicators separately entered into the regression equations.

According to these results, therefore, not only the age at which the teenager starts having sex and the number of sex partners but also the frequency with which he/she is engaged in sexual activities could significantly influence educational outcomes. It doesn’t seem that the measures of teenage sex we have analyzed are simply capturing the effects of the consequences of sex like child birth or related variables like teenage marriage since we mostly observe little changes in the coefficients when we control for these variables. In addition, there are some differences in the effects of teenage sex variables on the outcomes for boys and girls. While the underlying physiological, psychological or other reasons for gender differences in the effects of teenage sex on educational outcomes will need further investigation that is beyond the scope of this essay, it appears that the differences are non-trivial and survive our controls for teenage pregnancy that could potentially explain some of the differences.

However, the substantial differences between the OLS and the 2SLS estimates raise some questions as to whether the absolute magnitudes of the estimated coefficients are realistic. One possible reason for the attenuation of the OLS estimates could be that the downward bias in the estimates because of the under-reporting of teenage sex could be strongly offsetting the upward bias because of the other sources of endogeneity bias such as unobserved ability and motivation. Another possibility is that the complementary (positive) and substitution (negative) effects of teenage sex on schooling might be canceling out each other. On the other hand, the instruments based on parental and peer religiosity could only be identifying the negative effects of teenage sex. For example, if religious parents impose more discipline even in matters unrelated to sex, the religiosity instruments might be attributing the effects of these other elements of personal discipline on schooling to teenage sex. It could also be the case that children who belong to highly religious parents derive little positive stimulus from engaging in sexual activities perhaps because of what they have been taught about the evilness of premarital sex. In addition, the religiosity instruments may not be correcting for the reporting error bias in the estimates. Until all these issues are addressed in a future study perhaps using more detailed data for example on attitudes towards premarital sex and personal discipline, the results reported in this essay should be interpreted with caution.

Teenage Dating and High School Completion

We estimate the effects of alternative indicators of teenage involvement in dating activities on high school completion using the same procedure we followed in the case of teenage sex. Summaries of the OLS and 2SLS estimates for the effects of teenage dating are presented in table 17 below. The complete results out of which these summaries were extracted are reported in tables F9-F11 in appendix F. The marginal effects from probit models corresponding to these estimates are also reported in table F17 in the same appendix.

According to the first stage results for the 2SLS models reported in tables F3-F5 in appendix F, our instruments appear to be generally valid for age at first date both in terms of explanatory power and exogeneity as demonstrated by the statistics for joint and individual significance as well as the test for over-identification. The instruments are both individually and jointly significant in all of the first stage equations for age at first date. In addition the Sargan and Basmann chi-squared test for over-identification does not reject exogeneity of the instruments at 1% or 5% in all of the first stage equations for age at first date. On the other hand, the instruments turned out to be weak or invalid in some of the first stage equations for the teenage dating partners and frequency of dating. In models involving the number of dating partners, we do not reject exogeneity of the instruments but they are not jointly significant at conventional levels for the results reported in the last two columns of table17. The instruments are jointly significant in all of the first stage results for the frequency of teenage dating but we reject their exogeneity in all the cases. Therefore, the 2SLS results in this subsection are strictly valid only for age at first date and the rest of the results should be treated only as suggestive.

Table 17. Teenage Dating and High School Completion by Age 19-
Estimates from Linear Probability Models

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Dating Partners	-0.000**	-0.000	-0.000*	-0.045**	-0.043**	-0.040**

	(0.000)	(0.000)	(0.000)	(0.018)	(0.021)	(0.020)
Age at 1st date	0.017***	0.015***	0.014***	0.136***	0.132***	0.121***
	(0.002)	(0.002)	(0.002)	(0.025)	(0.029)	(0.028)
Total Dates/10	0.002***	0.001***	0.001***	0.058***	0.053***	0.048**
	(0.000)	(0.000)	(0.000)	(0.019)	(0.019)	(0.019)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	6026	6026	6026	6026	6026	6026

Robust standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the teenage dating indicators separately entered into the regression equations.

The effects of age at first date on the probability of high school completion reported in table 17 essentially mimic the corresponding estimates for age at first sex in terms of sign and statistical significance but the estimates for age at first date are somewhat larger in magnitude. In the model in which all the controls are included, delaying age at first date by an additional year increases the probability of high school completion by about 12% compared to 9.5% for age at first sex. The bigger effect of age at first date could be reflecting the fact that teenage dating on average starts much earlier than teenage sex (see table 13) and involves more partners and perhaps more time investment (or distraction from studies).

The estimates of the effects of the number of dating partners are somewhat smaller in magnitude and less significant than the corresponding estimates for the number of sex partners but the magnitudes of the 2SLS estimates of the former are not very reliable because of the poor explanatory power of the instruments in the first stage equations for the number of dating partners. The effect of the frequency of teenage dating appears to be unrelated to the frequency of teenage sex since the former is consistently positive while the latter is consistently negative both in the OLS and 2SLS results. If we were to believe our estimates for the effects of the frequency of teenage dating, the positive signs would imply that the purpose of dating is not limited to romantic activities but may rather work as a complement[48] to studying and hence may enhance schooling outcomes. In other words, the results may indicate that the positive effects of the complementary role outweigh the negative effects that arise because of the associated involvement in sexual activities and the required time investment. However, too much cannot be made of out of the 2SLS results for the frequency of teenage dating since the rejection of exogeneity of the instruments indicates that the models involving this variable may be mis-specified.

Table 18. Teenage Dating and High School Completion for Girls by Age 19-
Estimates from Linear Probability Models

		OLS			2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Dating Partners	0.000	0.000	0.000	-0.019*	-0.003	-0.005
	(0.000)	(0.000)	(0.000)	(0.011)	(0.009)	(0.008)
Age at 1st date	0.015***	0.012***	0.007**	0.110***	0.097***	0.068*
	(0.003)	(0.003)	(0.003)	(0.030)	(0.038)	(0.037)
Total Dates/10	0.002***	0.001***	0.001***	0.062*	0.052*	0.036
	(0.000)	(0.000)	(0.000)	(0.032)	(0.030)	(0.024)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	3054	3054	3054	3054	3054	3054

Robust standard errors in parentheses
***Significant at 1%; ** significant at 5%; * significant at 10%
Notes: Each of the teenage dating indicators separately entered into the regression equations.

[48] One possibility is that more attractive teens are likely to date more, all else the same, and "beauty" tends to be associated with a lot of positive outcomes in the labor market (e.g., Hamermesh and Biddle 1994; Fletcher, 2009) and probably schooling, if for no other reason than that employers (teachers) may reward attractive workers (students) better. However, there is so far little research that directly associates beauty and schooling outcomes. Mocan and Tekin (2009) find positive association between attractiveness and vocabulary test scores but not much is known about how beauty affects schooling outcomes in general.

As in the case of teenage sex, we try to look at whether the effects of teenage dating vary by gender by separately estimating the models for boys and girls. The summarized results for girls and boys are presented in tables 18 and 19, respectively. As in the case of age at first sex, age at first date has a larger effect on school outcomes for boys than girls. In the models where all the controls are included, delaying the initiation of teenage dating by a year increases the probability of high school graduation for boys by 14.4% compared to 6.8% for girls. In addition, we observe substantial reduction in the coefficient of age at first date for girls when we include controls for teenage pregnancy and child bearing while inclusion of teenage child birth leaves the coefficient nearly unchanged in the models for boys. This pattern in the estimates for the coefficient of age at first date is essentially similar to what we found in the case of age at first sex indicating that the two variables are mostly capturing the same information. While there are also non-negligible gender differences in the estimates for the coefficients of the number of teenage dating partners and frequency of dating, these results are not very informative due to the questionable validity of the instruments.

Table 19. Teenage Dating and High School Completion for Boys by Age 19-
Estimates from Linear Probability Models

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Dating Partners	-0.001**	-0.001**	-0.001**	-0.040*	-0.041	-0.041
	(0.000)	(0.000)	(0.000)	(0.024)	(0.027)	(0.027)
Age at 1st date	0.017***	0.015***	0.015***	0.151***	0.145***	0.144***
	(0.003)	(0.003)	(0.003)	(0.041)	(0.043)	(0.043)
Total Dates/10	0.001**	0.001*	0.001	0.040**	0.038**	0.039**
	(0.001)	(0.001)	(0.001)	(0.017)	(0.017)	(0.017)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation						

& Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	2972	2972	2972	2972	2972	2972

Robust standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the teenage dating indicators separately entered into the regression equations.

In general, although our finding about the importance of the age at which the teenager starts dating is informative, our results for teenage dating are mostly weaker than those we found for teenage sex. This is so because our instruments did not perform as well in explaining teenage dating behavior as they did in explaining teenage sex. Perhaps the constraints the religious values of parents and peers impose on the teenagers are less important in influencing their dating behavior in general than their involvement in sexual activities.

Teenage Dating, Sex and College Enrollment

While the main purpose of this essay was to examine the extent to which the failure to finish high school can be explained as a cumulative effect of teenage involvement in dating and sexual activities, we also tried to estimate models for college enrollment to see if involvement in dating and sexual activities as a teenager has any lasting effect beyond high-school completion. The covariates and estimation procedures for the college enrollment models are similar to those in the models for high school completion, but the college enrollment outcome is made conditional on high school completion. As previously stated GED recipients are included in the estimation sample as high school completers since GED is generally accepted for college admissions.

The summarized results for the effects of teenage sex and dating on college enrollment at or before age 20 are presented in table 20 below. The first stage results corresponding to the 2SLS estimates are reported in tables F12-F14 in appendix A. Although the explanatory power of the instruments is acceptable for all the indicators of teenage sex and dating as demonstrated by the joint significance of the instruments in the first stage results, their exogeneity is questionable in most of the models we tried for college enrollment. While we do not reject exogeneity of the instruments for the number of teenage sex partners and frequency of teenage dating on the basis of the Sargan and Basmann test for over-identification, we reject it at 5% or less for the other indictors of teenage sex and dating. This implies that our models for college enrollment are probably mis-specified, and the 2SLS results in this section should mostly be taken as suggestive at best.

Table 20. Linear Probability Models for College Enrollment at or Before Age 20 Conditional on High School Completion

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.005***	-0.004***	-0.004***	-0.068***	-0.068***	-0.063***
	(0.001)	(0.001)	(0.001)	(0.014)	(0.018)	(0.018)
Age at 1st sex	0.042***	0.035***	0.031***	0.152***	0.148***	0.141***
	(0.003)	(0.003)	(0.003)	(0.022)	(0.027)	(0.028)
Total sex /100	-0.016***	-0.009***	-0.007***	-0.190***	-0.180***	-0.168***
	(0.003)	(0.003)	(0.003)	(0.037)	(0.046)	(0.047)
Dating Partners	-0.001***	-0.001*	-0.001	-0.029**	-0.006	-0.004
	(0.000)	(0.000)	(0.000)	(0.013)	(0.011)	(0.011)
Age at 1st date	0.021***	0.017***	0.016***	0.167***	0.145***	0.131***
	(0.003)	(0.003)	(0.003)	(0.037)	(0.041)	(0.041)
Total Dates/10	0.001	-0.000	-0.000	0.098***	0.086***	0.081***
	(0.001)	(0.001)	(0.001)	(0.031)	(0.031)	(0.031)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	5208	5208	5208	5208	5208	5208

Standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the sex and dating experience indicators were separately entered into the regression equations.

While we do not want to make too much out of these results because of the questionable validity of the instruments, the fact that we obtain intuitively plausible signs with high level of statistical significance for all the coefficients of interest may be taken as indicative of the possible lasting effects of the teenage sex and dating on schooling outcomes. For example, the number of sex partners during the teenage years, for which we find the instruments to be valid, appears to have a large and highly significant negative effect on college enrollment-conditional on high school completion. In fact, the negative effect of the number of teenage sex partners on college enrollment is larger in magnitude than its effect on high school completion. The signs of the estimated coefficients for the frequency of teenage sex are also negative while the estimates for the coefficients of age at first sex are positive. The magnitudes of the estimated coefficients for both variables are larger in magnitude than the comparable estimates for high school completion. Although we could not directly verify the extent to which this is related to the effects of teenage sex on high school grades since data on the latter (collected only in 1999) are largely missing, it is possible that at least some of the observed effects could be through lower high school GPA.

Unlike the estimated coefficients for the number of teenage sex partners, the 2SLS estimates for the coefficients of the number of teenage dating partners are not only very small in magnitude but also statistically insignificant once the controls for teenage marriage, cohabitation and child birth are included. On the other hand, age at first date and frequency of teenage dating both have highly significant positive coefficients in the models for college enrollment just like their estimated coefficients in the models for high school completion. The magnitudes of the estimated coefficients for these two indicators are larger in the models for college enrollment than in their counterparts in the models for high school completion. However, the validity of the 2SLS estimates for the coefficients of number of dating partners and age at first date is questionable because of the rejection of exogeneity for the instruments in our models for college enrollment.

The gender differences we observe in the estimated coefficients of the indicators of teenage sex and dating in the models for college enrollment (reported in tables F15 and F16 in appendix F for girls and boys, respectively) are similar to the gender disparities we found in the estimated coefficients of these variables in the models for high school completion. For example, the estimated coefficient of the number of teenage sex partners is much larger for girls than boys, whereas age at first sex and frequency of teenage sex have larger coefficients in the models for boys than girls. Therefore, it appears that the same underlying reasons that drive the gender disparity in the effects of teenage sex and dating on high school completion explain the gender differences in the effects of these variables on college enrollment.

In general, while the veracity of some of our estimates could be questionable because of poor performance of our instruments in some of the models, the overall pattern of the results suggests that the choice teenagers make as to when to start dating and/or sex, how many dating/ sex partners to maintain, and how frequently to engage in sexual/dating activities could have significant effects not only on high school completion, but also the subsequent enrollment in a college. We observe only small changes in the estimated effects of the indicators of teenage sex and dating when we control for teenage pregnancy and child birth implying that the former are not just capturing the effects of the latter. The fact that parental and peer religiosity appears to constrain the teenagers' involvement in sex and dating (the effect being stronger on teenage sex) and the fact that the latter influence schooling outcomes implies that religious morality could be one way to influence schooling outcomes by imposing more discipline on the teenagers in terms of their involvement in sexual and dating activities. This poses an important policy dilemma (including some Constitutional issues) regarding the extent to which religious morality ought to be promoted in order to improve schooling outcomes. But it is important to note that such an outcome could be achieved through non-religious forms of parental discipline as long as these disciplinary measures succeed in convincing the teenager to delay the initiation of sex/dating, limit the number of sex/dating partners and frequency of sex/dating.

Conclusion

This essay uses data from the 1997 cohort of the National Longitudinal Survey of the Youth (NLSY97) to examine the extent to which high school completion (and to a limited extent college enrollment) are influenced by the choice teenagers make as to when to start dating and/or sex, how many dating and/or sex partners to maintain, and how frequently to engage in sexual and/or dating activities. We use indicators of parental and peer religiosity as instruments to address endogeneity of the teenager's involvement in sex and dating activities.

Our results indicate that the age at first sex, the number of sex partners and the frequency with which he/she is engaged in sexual activities as a teenager could significantly influence whether the child completes high school or not. For example, our preferred 2SLS estimates indicate that increasing ones sex partners between ages 15 to 17 by 1 reduces the probability of completing high school by as much as 4.6% on average while delaying initiation of sex by an additional year increases the probability of high school completion by 9.5%. It doesn’t seem that these indicators of teenage sex are simply capturing the effects of the consequences of sex such as child birth or related variables like teenage marriage since we typically observe small changes in the coefficients when we control for these variables. However, we find some differences in the effects of teenage sex variables on the outcomes for boys and girls. While the underlying physiological, psychological or other reasons for gender differences in the effects of teenage sex on educational outcomes will need further investigation that is beyond the scope of this essay, it appears that the differences are non-trivial and survive our controls for teenage pregnancy that could potentially explain some of the differences.

In the models for high school completion, our identifying instruments did not perform as well for teenage dating as they did for teenage sex. Perhaps the constraints the religious values of parents and peers impose on the teenagers are less important in influencing their dating behavior in general than their involvement in sexual activities. However, we find the instruments to be valid for age at first date that appears to mimic age at first sex in terms of sign and statistical significance of its estimated effect except that age at first date has a somewhat bigger effect on high school completion. In the model in which all the controls are included, we find that delaying age at first date by an additional year increases the probability of high school completion by about 12% compared to 9.5% for age at first sex. The bigger effect of age at first date could be reflecting the fact that teenage dating on average starts much earlier than teenage sex and involves more partners and perhaps more time investment.

Although most of our results for college enrollment are weaker than those for high school completion because of questionable validity of the instruments in most of our models for the former, the overall pattern of our estimates suggests that teenage sex and dating could have significant effects not only on high school completion, but also the subsequent enrollment in a college. The fact that parental and peer religiosity appears to constrain the teenagers' involvement in sex and dating (the effect being stronger on teenage sex) and the fact that the latter influence schooling outcomes implies that religious morality could be one (but not the only) way to influence schooling outcomes by imposing more discipline on the teenagers in terms of delaying initiation of sex and dating, limiting sex/dating partners and frequency of sex/dating. This poses an important policy dilemma (including some Constitutional issues) regarding the extent to which religious morality ought to be promoted in order to improve schooling outcomes.

APPENDICES

Appendix A. Additional Results for Essay 1

Table A1. Average Marginal Effects of Child's height-for-age z-scores on the Choice probabilities of Various Child Activities (Older Cohort)

Model	p(stud=1|x)	p(stud=1, work=0|x)	p(stud=1, work=1|x)	p(stud=0, work=1|x)	p(stud=0, work=0|x)	p(work=1|x)
Biprobit	-	0.001	0.037***	-0.032***	-0.006**	-
		(0.005)	(0.008)	(0.009)	(0.003)	
Biprobit two-stage	-	0.001	0.062	-0.052	-0.011	-
		(0.036)	(0.049)	(0.065)	(0.017)	
Mlogit	-	-0.005	0.043***	-0.035***	-0.002	-
		(0.007)	(0.009)	(0.010)	(0.004)	
Mlogit two-stage	-	0.026	0.044	-0.044	-0.026	-
		(0.044)	(0.064)	(0.064)	(0.020)	
Mlogit (Main act.)	0.036***	-	-	-0.030***	-0.007*	-
	(0.007)			(0.008)	(0.004)	
Mlogit two-stage (Main act.)	0.080	-	-	-0.064	-0.017	-
	(0.050)			(0.051)	(0.018)	
Probit	0.036***	-	-	-	-	0.008
	(0.007)					(0.007)
Probit, two-stage	0.079*	-	-	-	-	0.011
	(0.047)					(0.045)
Panel probit	0.034***	-	-	-	-	-
	(0.007)					
Panel prob. two- stage	0.055	-	-	-	-	-
	(0.049)					

*** p<0.01, ** p<0.05, *p<0.1

Notes: Table 4 in the main text and , tables A7, A8, and A9 in this appendix respectively present coefficient estimates for bivariate probit, multinomial logit, multinomial logit (Main activity), and Probit results from which these partial effects were obtained. The partial effects reported in this table are the averages of partial effects calculated at each value of the child's height-for-age z- scores and the standard errors were calculated by the delta method. All models included controls for community fixed effects, child age and sex, land and livestock owned, household size and number of siblings, education of father and mother, distance to primary school, age of mother and father, and sex of household head.

Table A2. Marginal effects (at the minimum value) of Child's height-for-age z-scores on the Choice Probabilities of Various Child Activities (Older Cohort)

Model	p(stud=1|x)	p(stud=1, work=0|x)	p(stud=1, work=1|x)	p(stud=0, work=1|x)	p(stud=0, work=0|x)	p(work=1|x)
Biprobit	0.031***	0.003	0.028***	-0.022***	-0.009	0.005
	(0.005)	(0.004)	(0.003)	(0.008)	(0.007)	(0.009)
Biprobit two-stage	0.037***	0.010	0.027*	-0.013	-0.023	0.014
	(0.003)	(0.018)	(0.016)	(0.072)	(0.071)	(0.072)
Mlogit	-	-0.000	0.023***	-0.023***	-0.000	-
		(0.001)	(0.002)	(0.002)	(0.000)	
Mlogit two-stage	-	0.000	0.023***	-0.023***	-0.000	-
		(0.000)	(0.002)	(0.003)	(0.000)	
Mlogit (Main act.)	0.025***	-	-	-0.020***	-0.005	-
	(0.002)			(0.005)	(0.005)	
Mlogit two-stage(Main act.)	0.024**	-	-	0.003	-0.027	-
	(0.011)			(0.077)	(0.078)	
Probit	0.026***	-	-	-	-	0.010
	(0.003)					(0.010)
Probit, two-stage	0.025*	-	-	-	-	0.015
	(0.013)					(0.079)
Panel probit	0.018***	-	-	-	-	-
	(0.002)					
Panel prob. two- stage	0.015	-	-	-	-	-
	(0.013)					

*** p<0.01, ** p<0.05, * p<0.1

Notes: Table 4 in the main text and tables A7, A8, and A9 in this appendix respectively present coefficient estimates for bivariate probit, multinomial logit, multinomial logit (Main activity), and Probit results from which these partial effects were obtained. The partial effects reported in this table were calculated at the sample minimum value of the child's height-for-age z-scores fixing other regressors at their mean values and the standard errors were calculated by the delta method. All models included controls for community fixed effects, child age and sex, land and livestock owned, household size and number of siblings, education of father and mother, distance to primary school, age of mother and father, and sex of household head.

Table A3. Marginal Effects (at maximum value) of Child's Height-for-age z-scores on the Choice probabilities of Various Child Activities (Older Cohort)

Model	p(stud=1|x)	p(stud=1, work=0|x)	p(stud=1, work=1|x)	p(stud=0, work=1|x)	p(stud=0, work=0|x)	p(work=1|x)
Biprobit	0.047***	-0.003	0.051***	-0.047***	-0.000	0.004
	(0.011)	(0.004)	(0.011)	(0.011)	(0.000)	(0.004)
Biprobit two-stage	0.067***	-0.006	0.073***	-0.068**	-0.000	0.006
	(0.029)	(0.010)	(0.026)	(0.029)	(0.000)	(0.009)
Mlogit	-	-0.000***	0.070***	-0.070***	-0.000**	-
		(0.000)	(0.012)	(0.012)	(0.000)	

Mlogit two-stage	-	0.003	0.070	-0.073	-0.000	-
		(0.018)	(0.081)	(0.073)	(0.000)	
Mlogit (main act.)	0.053***	-	-	-0.051***	-0.002***	-
	(0.010)			(0.010)	(0.000)	
Mlogit two-stage (main act.)	0.067**	-	-	-0.065**	-0.001	-
	(0.030)			(0.030)	(0.001)	
Probit	0.050***	-	-	-	-	0.005**
	(0.010)					(0.002)
Probit, two-stage	0.074***	-	-	-	-	0.005
	(0.017)					(0.006)
Panel probit	0.063**	-	-	-	-	-
	(0.015)					
Panel probit-two stage	0.091***	-	-	-	-	-
	(0.030)					

***p<0.01, ** p<0.05, * p<0.1

Notes: Table 4 in the main text and tables A7, A8, and A9 in this appendix respectively present coefficient estimates for bivariate probit, multinomial logit, multinomial logit (Main activity), Probit and Panel Probit results from which these partial effects were obtained. The partial effects reported in this table were calculated at the sample maximum value of the child's height-for-age z-scores fixing other regressors at their mean values and the standard errors were calculated by the delta method. All models included controls for community fixed effects, child age and sex, land and livestock owned, household size and number of siblings, education of father and mother, distance to primary school, age of mother and father, and sex of household head.

Table A4. Average Marginal Effects of Child's height-for-age z-scores on the Choice probabilities of various Child Activities (Younger Cohort)

Model	p(stud=1|x)	p(stud=1, work=0|x)	p(stud=1, work=1|x)	p(stud=0, work=1|x)	p(stud=0, work=0|x)	p(work=1|x)
Biprobit	-	0.005	0.021***	-0.017***	-0.009***	-
		(0.003)	(0.004)	(0.005)	(0.002)	
Biprobit two-stage	-	-0.004	0.080***	-0.034	-0.041***	-
		(0.019)	(0.022)	(0.027)	(0.014)	
Mlogit	-	0.010***	0.016***	-0.013***	-0.014***	-
		(0.003)	(0.005)	(0.006)	(0.003)	
Mlogit two-stage	-	0.026	0.047*	-0.007	-0.066***	-
		(0.024)	(0.028)	(0.031)	(0.018)	
Mlogit (main act.)	0.026***	-	-	-0.013***	-0.013***	-
	(0.004)			(0.005)	(0.003)	
Mlogit two-stage (main act.)	0.073**	-	-	-0.007	-0.066***	-
	(0.029)			(0.031)	(0.018)	
Probit	0.026***	-	-	-	-	0.007
	(0.005)					(0.015)

Probit, two-stage	0.073**	-	-	-	-	0.043
	(0.029)					(0.029)
Panel probit	0.029***	-	-	-	-	0.002
	(0.005)					(0.004)
Panel probit-two stage	0.082**	-	-	-	-	0.040
	(0.032)					(0.027)

***p<0.01, ** p<0.05, * p<0.1

Notes: Table 4 in the main text and , tables A7, A8, and A9 in this appendix respectively present coefficient estimates for bivariate probit, multinomial logit, multinomial logit (Main activity), Probit results from which these partial effects were obtained. The partial effects reported in this table are the averages of partial effects calculated at each value of the child's height-for-age z- scores and the standard errors were calculated by the delta method. All models included controls for community fixed effects, child age and sex, land and livestock owned, household size and number of siblings, education of father and mother, distance to primary school, age of mother and father, and sex of household head.

Table A5. Marginal effects (at the minimum value) of Child's height-for-age z-scores on the Choice Probabilities of Various Child Activities (Younger Cohort)

Model	p(stud=1|x)	p(stud=1, work=0|x)	p(stud=1, work=1|x)	p(stud=0, work=1|x)	p(stud=0, work=0|x)	p(work=1|x)
Biprobit	0.031***	0.007**	0.024***	-0.020***	-0.011***	0.004
	(0.005)	(0.003)	(0.003)	(0.006)	(0.004)	(0.005)
Biprobit two-stage	0.043**	0.029***	0.013	0.051	-0.094**	0.065*
	(0.011)	(0.008)	(0.011)	(0.045)	(0.046)	(0.039)
Mlogit	-	0.010***	0.022***	-0.018***	-0.014**	-
		(0.002)	(0.004)	(0.007)	(0.006)	
Mlogit two-stage	-	0.011	0.024	0.110*	-0.145	-
		(0.012)	(0.023)	(0.066)	(0.089)	
Mlogit (main act.)	0.033***	-	-	-0.021***	-0.012**	-
	(0.005)			(0.006)	(0.005)	
Mlogit two-stage (main act.)	0.038	-	-	0.110*	-0.147*	-
	(0.033)			(0.064)	(0.089)	
Probit	0.031***	-	-	-	-	0.002
	(0.005)					(0.005)
Probit, two-stage	0.043***	-	-	-	-	0.061
	(0.010)					(0.040)
Panel probit	0.036***	-	-	-	-	0.002
	(0.005)					(0.005)
Panel probit-two stage	0.035*	-	-	-	-	0.057
	(0.020)					(0.038)

*** p<0.01, ** p<0.05, * p<0.1

Notes: Table 4 in the main text and , tables A7, A8, and A9 in this appendix respectively present coefficient estimates for bivariate probit, multinomial logit, multinomial logit (Main activity), and Probit results from which these partial effects were obtained. The partial effects reported in this table were calculated at the sample minimum value of the child's height-for-age z-scores fixing other regressors at their mean values and the standard errors were calculated by the delta method. All models included controls for community fixed effects, child age and sex, land and livestock owned, household size and number of siblings, education of father and mother, distance to primary school, age of mother and father, and sex of household head.

Table A6. Marginal Effects (at maximum value) of Child's Height-for-age z-scores on the Choice probabilities of Various Child Activities (Younger Cohort)

Model	p(stud=1|x)	p(stud=1, work=0|x)	p(stud=1, work=1|x)	p(stud=0, work=1|x)	p(stud=0, work=0|x)	p(work=1|x)
Biprobit	0.025***	-0.002	0.026***	-0.023***	-0.002***	0.004
	(0.002)	(0.004)	(0.004)	(0.002)	(0.000)	(0.004)
Biprobit two-stage	0.006	-0.004	0.010	-0.006	-0.000	0.004
	(0.011)	(0.007)	(0.011)	(0.011)	(0.000)	(0.007)
Mlogit	-	0.010	0.012*	-0.022***	-0.001***	-
		(0.007)	(0.007)	(0.002)	(0.000)	
Mlogit two-stage	-	0.004	0.011	-0.015	-0.000	-
		(0.054)	(0.055)	(0.013)	(0.000)	
Mlogit (main act.)	0.021***	-	-	-0.021***	-0.001***	-
	(0.001)			(0.001)	(0.000)	
Mlogit two-stage (main act.)	0.013	-	-	-0.013	-0.000	-
	(0.012)			(0.012)	(0.000)	
Probit	0.025***	-	-	-	-	0.002
	(0.002)					(0.004)
Probit, two-stage	0.007	-	-	-	-	0.005
	(0.012)					(0.008)
Panel probit	0.026***	-	-	-	-	0.002
	(0.001)					(0.004)
Panel probit-two stage	0.002	-	-	-	-	0.006
	(0.006)					(0.008)

***p<0.01, ** p<0.05, * p<0.1

Notes: Table 4 in the main text and , tables A7, A8, and A9 in this appendix respectively present coefficient estimates for bivariate probit, multinomial logit, multinomial logit (Main activity), and Probit results from which these partial effects were obtained. The partial effects reported in this table were calculated at the sample maximum value of the child's height-for-age z-scores fixing other regressors at their mean values and the standard errors were calculated by the delta method. All models included controls for community fixed effects, child age and sex, land and livestock owned, household size and number of siblings, education of father and mother, distance to primary school, age of mother and father, and sex of household head.

Table A7. Multinomial Logit Estimates for Child Activities

	Older Cohort			Younger Cohort		
	I	II	III	IV	V	VI
	Mlogit	2-Stage Mlogit	Mlogit	2-Stage Mlogit	RE Mlogit	2-Stage RE Mlogit
Student						
Child's Height-for-age	-0.019	0.379	0.138***	0.296	0.151***	0.285
	(0.096)	(0.637)	(0.038)	(0.240)	(0.041)	(0.257)
Sex	0.395*	0.469*	0.438***	0.477***	0.480***	0.518***
	(0.232)	(0.259)	(0.149)	(0.158)	(0.161)	(0.173)
Age	-0.088	-0.044	0.223***	0.227***	0.168***	0.162***
	(0.074)	(0.103)	(0.044)	(0.044)	(0.036)	(0.037)
Agri. Land Area Owned	-0.019	-0.027	-0.157*	-0.159*	-0.164*	-0.165*
	(0.109)	(0.109)	(0.091)	(0.091)	(0.093)	(0.093)
Livestock Units Owned	0.005	0.006	0.033	0.034	0.032	0.034
	(0.028)	(0.028)	(0.022)	(0.022)	(0.023)	(0.023)
Father's Education	1.135*	0.992	0.825***	0.777***	0.900***	0.858***
	(0.600)	(0.690)	(0.213)	(0.222)	(0.234)	(0.246)
Mother's Education	0.026	0.050	0.489	0.542	0.541	0.595*
	(1.139)	(1.148)	(0.324)	(0.334)	(0.347)	(0.362)
Distance to Primary School	0.031	0.030	-0.064	-0.063	0.009	-0.062
	(0.080)	(0.080)	(0.040)	(0.040)	(0.045)	(0.041)
Household Size	-0.006	0.001	-0.057	-0.061	-0.028	-0.025
	(0.079)	(0.081)	(0.042)	(0.042)	(0.040)	(0.041)
Number of Siblings	0.068	0.055	0.123**	0.129**	0.099*	0.103*
	(0.076)	(0.080)	(0.051)	(0.051)	(0.053)	(0.054)
Age of Father	0.005	0.003	-0.012	-0.013	-0.015	-0.016
	(0.019)	(0.019)	(0.011)	(0.011)	(0.012)	(0.012)
Age of Mother	0.001	-0.007	-0.001	-0.002	-0.001	-0.002
	(0.025)	(0.029)	(0.014)	(0.014)	(0.015)	(0.015)
Sex of Household Head	-0.382	-0.394	0.111	0.117	0.147	0.119
	(0.426)	(0.428)	(0.231)	(0.230)	(0.248)	(0.246)
Year=2004			-0.723***	-0.738***		
			(0.261)	(0.261)		
Resid. from Height-for-age eqn.		-0.408		-0.162		-0.136
		(0.644)		(0.245)		(0.262)
Constant	-1.909	-1.152	-3.409***	-2.969***	-3.287***	-2.652***
	(1.814)	(2.206)	(0.639)	(0.937)	(0.822)	(1.009)
Student and Working						
Child's Height-for-age z-scores	0.287***	0.299	0.123***	0.300*	0.140***	0.291
	(0.063)	(0.440)	(0.034)	(0.180)	(0.037)	(0.203)
Sex	0.675***	0.677***	0.483***	0.528***	0.539***	0.585***
	(0.163)	(0.176)	(0.124)	(0.132)	(0.139)	(0.148)
Age	0.121**	0.123*	0.253***	0.256***	0.366***	0.357***
	(0.049)	(0.065)	(0.036)	(0.036)	(0.028)	(0.029)
Agri. Land Area Owned	-0.013	-0.013	0.033	0.032	0.018	0.021
	(0.048)	(0.049)	(0.050)	(0.050)	(0.053)	(0.054)
Livestock Units Owned	0.019	0.019	0.035**	0.036***	0.048***	0.051***
	(0.019)	(0.018)	(0.014)	(0.014)	(0.016)	(0.016)
Father's Educ.-at Least Primary	1.400***	1.395***	0.689***	0.633***	0.823***	0.775***
	(0.332)	(0.369)	(0.220)	(0.222)	(0.245)	(0.251)

Mother's Educ.-at Least Primary	0.007	0.009	0.806***	0.865***	0.881***	0.952***
	(0.909)	(0.912)	(0.277)	(0.278)	(0.304)	(0.309)
Distance to Primary School	0.074	0.074	-0.079**	-0.079**	0.021	-0.117***
	(0.052)	(0.052)	(0.036)	(0.036)	(0.035)	(0.040)
Household Size	0.039	0.040	-0.041	-0.046	-0.081**	-0.078**
	(0.040)	(0.041)	(0.031)	(0.031)	(0.033)	(0.034)
Number of Siblings	-0.000	-0.001	0.058	0.064	0.089*	0.094**
	(0.043)	(0.047)	(0.041)	(0.041)	(0.047)	(0.047)
Age of Father	-0.013	-0.013	-0.010	-0.011	-0.013	-0.014
	(0.010)	(0.010)	(0.008)	(0.008)	(0.009)	(0.010)
Age of Mother	0.012	0.012	0.010	0.008	0.012	0.010
	(0.011)	(0.014)	(0.011)	(0.011)	(0.012)	(0.012)
Sex of Household Head	0.140	0.140	0.258	0.266	0.319	0.278
	(0.253)	(0.254)	(0.177)	(0.177)	(0.202)	(0.200)
Year==2004			0.579***	0.566***		
			(0.204)	(0.204)		
Residuals from 1st stage.		-0.012		-0.181		-0.154
		(0.450)		(0.180)		(0.203)
Constant	-3.499***	-3.482**	-3.662***	-3.155***	-4.787***	-3.862***
	(1.122)	(1.519)	(0.535)	(0.753)	0.688	(0.835)
Neither Student nor Working						
Child's Height-for-age	-0.071	-1.512	-0.161***	-0.905***	-0.164***	-0.960***
	(0.254)	(1.249)	(0.045)	(0.280)	(0.050)	(0.294)
Sex	1.583***	1.335**	-0.004	-0.185	0.049	-0.147
	(0.554)	(0.551)	(0.178)	(0.180)	(0.194)	(0.197)
Age	0.036	-0.086	-0.794***	-0.810***	-0.696***	-0.704***
	(0.157)	(0.171)	(0.065)	(0.067)	(0.060)	(0.061)
Agri. Land Area Owned	0.086	0.118	-0.188**	-0.178*	-0.177*	-0.177*
	(0.108)	(0.110)	(0.096)	(0.095)	(0.105)	(0.103)
Livestock Units Owned	-0.115*	-0.116*	-0.055*	-0.064**	-0.052	-0.058*
	(0.065)	(0.065)	(0.030)	(0.031)	(0.033)	(0.033)
Father's Education	-33.721***	-34.250***	0.239	0.492*	0.321	0.583*
	(0.623)	(0.710)	(0.273)	(0.281)	(0.291)	(0.302)
Mother's Education	-32.934***	-34.189***	-0.471	-0.713	-0.402	-0.670
	(0.829)	(0.858)	(0.500)	(0.507)	(0.522)	(0.530)
Distance to Primary School	0.066	0.062	0.001	0.010	-0.045	-0.054
	(0.127)	(0.123)	(0.061)	(0.062)	(0.052)	(0.064)
Household Size	0.207**	0.178**	-0.005	0.016	-0.025	-0.001
	(0.100)	(0.087)	(0.040)	(0.042)	(0.044)	(0.046)
Number of Siblings	-0.049	0.004	0.058	0.031	0.083	0.054
	(0.111)	(0.114)	(0.061)	(0.062)	(0.068)	(0.068)
Age of Father	-0.019	-0.013	-0.017	-0.014	-0.017	-0.014
	(0.048)	(0.047)	(0.013)	(0.013)	(0.013)	(0.014)
Age of Mother	-0.007	0.023	0.003	0.009	0.001	0.007
	(0.045)	(0.052)	(0.015)	(0.015)	(0.016)	(0.017)
Sex of Household Head	-0.808	-0.747	-0.059	-0.044	-0.056	-0.040
	(0.595)	(0.567)	(0.258)	(0.257)	(0.279)	(0.277)
Year=2004			1.205***	1.299***		
			(0.398)	(0.408)		
Residuals from the 1st stage		1.481		0.762***		0.812***

		(1.297)		(0.287)		(0.299)
Constant	-4.576	-7.878*	5.296***	3.149***	5.214***	2.464**
	(3.186)	(4.373)	(0.740)	(1.027)	(0.921)	(1.129)
St. dev of individual random effects[a]					-0.492*	0.528**
					(0.281)	(0.266)
St. dev of family random effects					0.810***	0.827***
					(0.128)	(0.129)
Number of individuals					1263	1263
Number of families					781	781
Observations	1116	1116	2241	2241	2241	2241

*** p<0.01, ** p<0.05, * p<0.1

Cluster-robust standard errors in parentheses

Notes: Dummies representing exposure to a big drought in 1984 at 1st, 2nd and 3rd years are used as identifying instruments for child's height-for-age in equation (II) . Dummies for substantial rain deficit and rain surplus at 1st, 2nd and 3rd years are used as instruments in IV and VI. Mother's height and father's height were also included in all first stage equations to control for genetic variations in height. Site dummies were included in all equations to control for community fixed effects. The excluded child activity in all equations was 'work only'.

[a] When the estimated parameter is negative its absolute value is interpreted as the standard deviation (Rabe-Hesketh, Skrondal, and Pickles, 2004).

Table A8. Multinomial Logit Estimates for Main Child Activities

	Older Cohort				Younger Cohort			
	Pooled Panel		Panel		Pooled Panel		Panel	
	I	II	III	IV	V	VI	VII	VIII
		2-Stage	RE	2-Stage RE		2-Stage	RE	2-Stage RE
	Mlogit	Mlogit	Mlogit	Mlogit	Mlogit	Mlogit	Mlogit	Mlogit
Student								
Child's Height-for -age	0.210***	0.463	0.206***	0.321	0.128***	0.300*	0.144***	0.293
	(0.045)	(0.303)	(0.058)	(0.345)	(0.032)	(0.166)	(0.036)	(0.188)
Sex	0.778***	0.839***	1.094***	1.121***	0.466***	0.509***	0.517***	0.557***
	(0.122)	(0.141)	(0.174	(0.193)	(0.115)	(0.122)	(0.130)	(0.138)
Age	0.106***	0.130***	0.166***	0.178***	0.245***	0.249***	0.303***	0.305***
	(0.034)	(0.045)	(0.044)	(0.057)	(0.034)	(0.034)	(0.026)	(0.026)
Agri. Land Area Owned	0.013	0.012	0.032	0.032	-0.009	-0.010	-0.018	-0.019
	(0.043)	(0.043)	(0.061)	(0.061)	(0.038)	(0.039)	(0.043)	(0.043)
Livestock Units Owned	0.009	0.009	0.014	0.014	0.035***	0.036***	0.044***	0.045***
	(0.015)	(0.015)	(0.023)	(0.023)	(0.013)	(0.013)	(0.015)	(0.015)
Father's Education	1.592***	1.502***	2.110***	2.070***	0.740***	0.686***	0.845***	0.802***
	(0.281)	(0.298)	(0.405)	(0.418)	(0.192)	(0.194)	(0.217)	(0.222)
Mother's Education	0.112	0.119	0.036	0.040	0.705***	0.763***	0.776***	0.828***
	(0.709)	(0.703)	(0.887)	(0.885)	(0.250)	(0.252)	(0.280)	(0.283)

Distance to Primary School	0.066*	0.065	0.100*	0.099*	-0.051*	-0.051*	0.018	0.018
	(0.039)	(0.040)	(0.052)	(0.052)	(0.027)	(0.027)	(0.033)	(0.033)
Household Size	0.017	0.025	0.009	0.013	-0.049*	-0.054*	-.062**	-.066**
	(0.032)	(0.033)	(0.046)	(0.046)	(0.028)	(0.028)	(0.030)	(0.031)
Number of Siblings	0.019	0.005	0.051	0.045	0.078**	0.084**	0.087**	0.093**
	(0.035)	(0.038)	(0.051)	(0.053)	(0.036)	(0.036)	(0.041)	(0.042)
Age of Father	-0.004	-0.005	-0.006	-0.006	-0.011	-0.012	-0.013	-0.014
	(0.008)	(0.008)	(0.011)	(0.011)	(0.008)	(0.008)	(0.009)	(0.009)
Age of Mother	0.006	-0.000	0.005	0.002	0.007	0.005	0.008	0.007
	(0.009)	(0.011)	(0.013)	(0.015)	(0.010)	(0.010)	(0.011)	(0.011)
Sex of Household Head	0.058	0.062	-0.039	-0.037	0.221	0.228	0.266	0.273
	(0.181)	(0.181)	(0.247)	(0.247)	(0.164)	(0.163)	(0.188)	(0.188)
Year=1995	0.486***	0.511***						
	(0.098)	(0.102)						
Year=2004					0.178	0.164		
					(0.181)	(0.180)		
Residuals from 1st stage		-0.258		-0.116		-0.176		-0.153
		(0.306)		(0.345)		(0.166)		(0.188)
Constant	-3.53***	-3.00***	-4.81***	-4.57***	-2.92***	-2.43***	-3.64***	-3.21***
	(0.796)	(1.021)	(1.079)	(1.331)	(0.485)	(0.690)	(0.626)	(0.857)
Neither Student nor Working								
Child's Height-for -age	-0.115	-0.313	-0.107	-0.149	-.162***	-.902***	-.166***	-.938***
	(0.093)	(0.470)	(0.108)	(0.502)	(0.045)	(0.280)	(0.050)	(0.283)
Sex	0.441**	0.393*	0.664***	0.653**	-0.004	-0.183	0.046	-0.142
	(0.219)	(0.234)	(0.252)	(0.269)	(0.178)	(0.180)	(0.193)	(0.197)
Age	-.296***	-.318***	-.379***	-.385***	-.793***	-.809***	-.688***	-.692***
	(0.067)	(0.081)	(0.088)	(0.102)	(0.065)	(0.067)	(0.060)	(0.060)
Agri. Land Area Owned	0.069	0.070	0.027	0.027	-0.178*	-0.168*	-0.160	-0.149
	(0.094)	(0.094)	(0.105)	(0.105)	(0.095)	(0.094)	(0.103)	(0.102)
Livestock Units Owned	-0.029	-0.028	-0.036	-0.036	-0.055*	-.063**	-0.051	-0.058*
	(0.026)	(0.026)	(0.031)	(0.031)	(0.030)	(0.031)	(0.033)	(0.033)
Father's Education	-0.109	-0.046	0.384	0.395	0.231	0.482*	0.301	0.562*
	(0.449)	(0.467)	(0.523)	(0.546)	(0.273)	(0.280)	(0.291)	(0.299)
Mother's Education	-0.831	-0.805	-1.431	-1.413	-0.459	-0.700	-0.387	-0.637
	(0.819)	(0.823)	(1.077)	(1.078)	(0.500)	(0.507)	(0.522)	(0.530)
Distance to Primary School	0.178*	0.181*	0.203**	0.203**	0.003	0.012	-0.045	-0.038
	(0.096)	(0.096)	(0.103)	(0.103)	(0.061)	(0.062)	(0.051)	(0.050)
Household Size	0.122**	0.115**	0.159**	0.157**	-0.005	0.015	-0.027	-0.008

	(0.052)	(0.054)	(0.064)	(0.067)	(0.040)	(0.042)	(0.044)	(0.045)
Number of Siblings	-0.106*	-0.094	-0.127*	-0.124*	0.055	0.029	0.080	0.054
	(0.055)	(0.061)	(0.068)	(0.074)	(0.061)	(0.062)	(0.068)	(0.068)
Age of Father	-0.007	-0.007	-0.008	-0.008	-0.017	-0.014	-0.016	-0.013
	(0.016)	(0.016)	(0.017)	(0.017)	(0.013)	(0.013)	(0.013)	(0.014)
Age of Mother	-0.010	-0.005	-0.006	-0.005	0.003	0.009	0.001	0.007
	(0.018)	(0.019)	(0.020)	(0.021)	(0.015)	(0.015)	(0.016)	(0.017)
Sex of Household Head	-.976***	-.980***	-.957***	-.957***	-0.052	-0.037	-0.048	-0.032
	(0.285)	(0.286)	(0.326)	(0.326)	(0.257)	(0.256)	(0.277)	(0.275)
Year=1995	-1.40***	-1.41***						
	(0.291)	(0.291)						
Year=2004					1.223***	1.315***		
					(0.398)	(0.408)		
Residuals from 1st stage.		0.204		0.044		0.758***		0.790***
		(0.463)		(0.492)		(0.286)		(0.290)
Constant	-0.064	-0.501	-0.474	-0.557	5.274***	3.138***	5.131***	2.762**
	(1.793)	(2.066)	(2.008)	(2.190)	(0.740)	(1.024)	(0.914)	(1.169)
St. dev of ind. Rand. effects			1.220***	1.222***			0.472*	0.491*
			(0.211)	(0.211)			(0.287)	(0.279)
St. dev of family rand. effects			1.137***	1.134***			0.801***	0.802***
			(0.162)	(0.163)			(0.127)	(0.128)
No. of Individuals			1358	1358			1263	1263
No. of families			840	840			781	781
Observations	2348	2348	2348	2348	2241	2241	2241	2241

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Robust standard errors in parentheses

Notes: Dummies representing exposure to a big drought in 1984 at 1st, 2nd and 3rd years are used as identifying instruments for child's height-for-age in equation II and IV. Dummies for substantial rain deficit and rain surplus at 1st, 2nd and 3rd years are used as instruments in VI and VIII. Mother's height and father's height were also included in all first stage equations to control for genetic variation in heights. Site dummies were included in all equations to control for community fixed effects. The excluded child activity in all equations was 'work'.

Table A9. Probit Models for Child Schooling

	Pooled Panel		Panel		Pooled Panel		Panel	
	I	II	III	IV	V	VI	VII	VII
	Probit	2-Stage Probit	RE Probit	2-Stage RE Probit	Probit	2-Stage Probit	RE Probit	2-Stage RE Probit
Child's Height-for-age	0.125***	0.276*	0.159***	0.253	0.085***	0.242**	0.105***	0.299**
	(0.025)	(0.166)	(0.035)	(0.228)	(0.018)	(0.097)	(0.020)	(0.120)
Sex	0.441***	0.477***	0.664***	0.686***	0.276***	0.314***	0.339***	0.385***
	(0.068)	(0.078)	(0.107)	(0.120)	(0.066)	(0.070)	(0.080)	(0.085)
Age	0.072***	0.087***	0.152***	0.161***	0.201***	0.204***	0.251***	0.252***
	(0.019)	(0.025)	(0.030)	(0.038)	(0.018)	(0.018)	(0.018)	(0.018)
Agri. Land Area Owned	0.003	0.002	0.027	0.027	-0.001	-0.002	-0.001	-0.002
	(0.025)	(0.025)	(0.033)	(0.033)	(0.024)	(0.024)	(0.035)	(0.035)
Livestock Units Owned	0.007	0.006	0.009	0.009	0.022***	0.023***	0.028***	0.029***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.007)	(0.008)	(0.009)	(0.009)
Father's Education	0.951***	0.898***	1.453***	1.420***	0.415***	0.366***	0.500***	0.440***
	(0.154)	(0.163)	(0.213)	(0.227)	(0.106)	(0.108)	(0.122)	(0.126)
Mother's Education	0.055	0.059	0.008	0.011	0.443***	0.495***	0.524***	0.587***
	(0.376)	(0.374)	(0.388)	(0.388)	(0.144)	(0.145)	(0.195)	(0.199)
Distance to Primary School	0.034	0.033	0.042	0.041	-0.024	-0.024	-0.034	-0.034
	(0.023)	(0.024)	(0.032)	(0.032)	(0.015)	(0.015)	(0.022)	(0.022)
Household Size	0.007	0.012	-0.006	-0.003	-0.025	-0.030*	-0.029	-0.033*
	(0.018)	(0.019)	(0.026)	(0.027)	(0.016)	(0.016)	(0.018)	(0.018)
Number of Siblings	0.014	0.005	0.041	0.036	0.048**	0.053**	0.057**	0.063**
	(0.020)	(0.022)	(0.028)	(0.030)	(0.021)	(0.021)	(0.025)	(0.026)
Age of Father	-0.002	-0.003	-0.002	-0.002	-0.005	-0.006	-0.007	-0.008
	(0.005)	(0.005)	(0.007)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)
Age of Mother	0.003	0.000	0.005	0.003	0.004	0.002	0.005	0.003
	(0.005)	(0.007)	(0.008)	(0.009)	(0.006)	(0.006)	(0.007)	(0.007)
Sex of Household Head	0.070	0.073	0.040	0.043	0.124	0.129	0.133	0.138
	(0.103)	(0.103)	(0.145)	(0.145)	(0.095)	(0.094)	(0.103)	(0.103)
Year=1995	0.331***	0.345***						
	(0.056)	(0.058)						
Year=2004					0.031	0.018		
					(0.105)	(0.105)		
Residuals from 1st stage.		-0.154		-0.096		-0.161*		-0.199*
		(0.168)		(0.230)		(0.097)		(0.121)
Constant	-2.23***	-1.91***	-3.61***	-3.41***	-2.46***	-2.00***	-3.03***	-2.45***

	(0.456)	(0.577)	(0.698)	(0.842)	(0.270)	(0.393)	(0.316)	(0.462)
lnsig2u			0.224	0.223			-0.69***	-0.70***
			(0.190)	(0.190)			(0.264)	(0.265)
No. of Individuals			1358	1358			1263	1263
Observations	2348	2348	2348	2348	2241	2241	2241	2241

*** p<0.01, ** p<0.05, * p<0.1

Robust standard errors in parentheses

Notes: Dummies representing exposure to a big drought in 1984 at 1st, 2nd and 3rd years are used as identifying instruments for child's height-for-age in (II) &(IV) . Dummies for substantial rain deficit and rain surplus at 1st, 2nd and 3rd years are used as instruments in (VI)&(VIII). Mother's height and father's height were also included in all first stage equations to control for genetic variations in height. Site dummies were included in all the equations to control for community fixed effects.

Table A10 Semi-Nonparametric Bivariate Estimates for Child Schooling and Work

	Older Cohort		Younger Cohort	
	I	II	III	IV
Variables	SNP	2-Stage SNP	SNP	2-Stage SNP
Stud				
Child's Height-for-age z-scores	0.169***	0.334	0.097***	0.304***
	(0.042)	(0.222)	(0.022)	(0.100)
Sex	0.468***	0.499***	0.353***	0.410***
	(0.105)	(0.113)	(0.089)	(0.103)
Age	0.046	0.066	0.223***	0.230***
	(0.029)	(0.043)	(0.028)	(0.027)
Total Agricultural Land Area Owned	-0.031	-0.035	0.010	0.009
	(0.038)	(0.038)	(0.034)	(0.034)
Total Tropical Livestock Units Owned	0.012	0.013	0.027***	0.028***
	(0.012)	(0.012)	(0.010)	(0.010)
Father's Education-at Least Primary	1.129***	1.076***	0.484***	0.428***
	(0.229)	(0.237)	(0.131)	(0.133)
Mother's Education-at Least Primary	-0.045	-0.016	0.494**	0.562***
	(0.517)	(0.514)	(0.198)	(0.193)
Distance to the Nearest Primary School	0.031	0.036	-0.071***	-0.069***
	(0.033)	(0.035)	(0.022)	(0.022)
Household Size	0.034	0.038	-0.030	-0.035
	(0.031)	(0.031)	(0.023)	(0.023)
Number of Siblings	0.001	-0.005	0.053*	0.061**
	(0.035)	(0.036)	(0.029)	(0.029)
Age of Father	-0.006	-0.006	-0.009	-0.009
	(0.008)	(0.008)	(0.007)	(0.007)
Age of Mother	0.005	0.002	0.005	0.004
	(0.009)	(0.009)	(0.008)	(0.008)
Sex of Household Head	-0.032	-0.028	0.125	0.135
	(0.168)	(0.167)	(0.129)	(0.129)
year04			0.164	0.139
			(0.189)	(0.197)
Child's Height-for-age z-scores				

Resid. from Height-for-age eqn.		-0.168		-0.212**
		(0.227)		(0.099)
work				
Child's Height-for-age z-scores	0.032	-0.369	0.030	0.149
	(0.046)	(0.269)	(0.022)	(0.093)
Sex	-0.321**	-0.390***	-0.072	-0.050
	(0.135)	(0.144)	(0.087)	(0.090)
Age	0.131***	0.061	0.250***	0.248***
	(0.033)	(0.051)	(0.029)	(0.033)
Total Agricultural Land Area Owned	0.008	0.021	0.105**	0.105**
	(0.050)	(0.050)	(0.048)	(0.049)
Total Tropical Livestock Units Owned	0.007	0.006	-0.000	0.001
	(0.022)	(0.022)	(0.013)	(0.013)
Father's Education-at Least Primary	-0.204	-0.104	-0.193	-0.237*
	(0.287)	(0.312)	(0.125)	(0.126)
Mother's Education-at Least Primary	0.489	0.437	0.181	0.210
	(0.511)	(0.518)	(0.204)	(0.208)
Distance to the Nearest Primary School	0.065**	0.042	0.071**	0.068**
	(0.031)	(0.034)	(0.031)	(0.031)
Household Size	-0.021	-0.036	0.016	0.012
	(0.036)	(0.037)	(0.023)	(0.023)
Number of Siblings	-0.009	0.008	-0.041	-0.037
	(0.037)	(0.040)	(0.030)	(0.030)
Age of Father	0.001	0.002	0.011**	0.010*
	(0.010)	(0.010)	(0.006)	(0.006)
Age of Mother	0.009	0.014	0.004	0.002
	(0.012)	(0.012)	(0.008)	(0.008)
Sex of Household Head	0.417**	0.387**	0.106	0.097
	(0.189)	(0.190)	(0.135)	(0.138)
year04			-0.091	-0.086
			(0.210)	(0.221)
g_1_1	-0.685***	-0.684***	-0.066	-0.057
	(0.042)	(0.042)	(0.074)	(0.088)
g_2_1	0.058**	0.053**	0.270***	0.268***
	(0.023)	(0.023)	(0.042)	(0.042)
g_1_2			-0.225***	-0.225***
			(0.023)	(0.024)
g_2_2			-0.070***	-0.069***
			(0.017)	(0.014)
Standard Deviation (s)	1.281	1.281	1.301	1.300
Standard Deviation (w)	1.279	1.279	1.342	1.340
Variance (s)	1.642	1.640	1.692	1.690
Variance(w)	1.636	1.635	1.801	1.796
Skewness(s)	-0.072	-0.066	-0.397	-0.395
Skewness(w)	-0.005	-0.005	0.331	0.329
Kurtosis(s)	2.550	2.551	3.127	3.124
Kurtosis(w)	2.527	2.531	3.055	3.053
rho	-0.557	-0.559	-0.484	-0.482
Observations	1116	1116	2241	2241

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: The bivariate binary-choice models were estimated through the semi-nonparametric estimators of Gallant and Nychka (1987). The unknown density of the latent regression errors is approximated by a Hermite polynomial expansion of order 2 for and 1 for the schooling and work equations for the older cohort and order 2 in both equations for the younger cohort. Convergence was hard to obtain with higher order polynomials because of the non-concavity of the log pseudolikelihood function.

Appendix B. Robustness to Missing Rainfall Data

Table B1. Missing Monthly Rainfall Records for the 8 Critical Years in the Analysis of the Younger Cohort

	Year							
Site	1988	1989	1990	1991	1992	1993	1994	1995
Adado	3	0	0	5	0	3	0	0
Adele Keke	0	0	0	0	1	12	12	12
Aze Deboa	0	0	0	0	0	0	0	1
Debre berhan	2	0	0	1	0	3	1	1
Dinki	1	0	0	0	0	0	0	0
Doma'a	0	0	0	0	0	0	1	0
Gara Godo	4	1	1	1	2	2	0	0
Geblen	12	12	12	12	5	0	0	0
Haresaw	8	12	12	12	4	0	0	3
Imdibir	0	0	0	0	1	0	12	4
Koro degaga	0	0	0	4	4	2	0	0
Shumsha	0	5	12	12	6	1	0	1
Sirbana Gudeti	0	0	1	4	0	0	1	0
Tirufe Kecheme	0	0	0	1	0	0	1	0
Yetmen	0	0	3	0	0	0	0	0

Table B2. Availability of Rainfall Data During the Main Rainy Season

	Year							
Site	1988	1989	1990	1991	1992	1993	1994	1995
Adado	A	A	A	A	A	A	A	A
Adele Keke	A	A	A	A	A	NA	NA	NA
Aze Deboa	A	A	A	A	A	A	A	A
Debre berhan	A	A	A	A	A	A	A	A
Dinki	A	A	A	A	A	A	A	A
Doma'a	A	A	A	A	A	A	A	A
Gara Godo	A	A	A	A	A	A	A	A
Geblen	NA	NA	NA	NA	A	A	A	A
Haresaw	NA	NA	NA	NA	A	A	A	A
Imdibir	A	A	A	A	A	A	NA	A
Koro degaga	A	A	A	A	A	A	A	A
Shumsha	A	A	NA	NA	A	A	A	A
Sirbana Gudeti	A	A	A	A	A	A	A	A
Tirufe Kecheme	A	A	A	A	A	A	A	A
Yetmen	A	A	A	A	A	A	A	A

Notes: A stands for "data available" and NA stands for "data not available".

Table B3. Two-Stage Bivariate Probit Results-Successively Adjusting the Estimation Sample for the Major Missing Rainfall Records

VARIABLES	(I)	(II)	(III)	(IV)
Student				
Child's Height-for-age z-scores	0.251***	0.259***	0.233**	0.226**
	(0.095)	(0.095)	(0.101)	(0.105)
Sex	0.324***	0.326***	0.315***	0.314***
	(0.070)	(0.070)	(0.072)	(0.072)
Age	0.199***	0.199***	0.200***	0.210***
	(0.018)	(0.019)	(0.020)	(0.021)
Total Agricultural Land Area Owned	-0.002	-0.007	-0.006	-0.003
	(0.024)	(0.023)	(0.024)	(0.023)
Total Tropical Livestock Units Owned	0.024***	0.024***	0.023***	0.022***
	(0.008)	(0.008)	(0.008)	(0.008)
Father's Education-at Least Primary	0.349***	0.362***	0.377***	0.370***
	(0.107)	(0.107)	(0.109)	(0.114)
Mother's Education-at Least Primary	0.472***	0.489***	0.494***	0.502***
	(0.143)	(0.144)	(0.148)	(0.153)
Distance to the Nearest Primary School	-0.027*	-0.031**	-0.027*	-0.026
	(0.016)	(0.016)	(0.016)	(0.016)
Household Size	-0.030*	-0.027*	-0.030*	-0.028
	(0.016)	(0.017)	(0.017)	(0.017)
Number of Siblings	0.052**	0.054**	0.056**	0.061***
	(0.021)	(0.022)	(0.022)	(0.023)
Age of Father	-0.006	-0.007	-0.007	-0.005
	(0.005)	(0.005)	(0.005)	(0.005)
Age of Mother	0.001	0.001	0.003	0.002
	(0.006)	(0.006)	(0.006)	(0.007)
Sex of Household Head	0.113	0.079	0.077	0.079
	(0.094)	(0.096)	(0.099)	(0.105)
Resid. from Height-for-age eqn.	-0.169*	-0.175*	-0.150	-0.141
	(0.095)	(0.095)	(0.100)	(0.105)
Year=2004	0.043	0.065	0.069	0.045
	(0.105)	(0.111)	(0.112)	(0.115)
Constant	-1.892***	-1.965***	-2.059***	-2.229***
	(0.390)	(0.389)	(0.407)	(0.416)
Work				
Child's Height-for-age z-scores	0.162	0.176*	0.189*	0.174
	(0.103)	(0.102)	(0.106)	(0.115)
Sex	-0.035	-0.068	-0.098	-0.119*
	(0.065)	(0.066)	(0.068)	(0.069)
Age	0.141***	0.148***	0.140***	0.140***
	(0.018)	(0.019)	(0.020)	(0.021)
Total Agricultural Land Area Owned	0.081**	0.080**	0.076**	0.073**
	(0.037)	(0.037)	(0.037)	(0.037)
Total Tropical Livestock Units Owned	0.002	0.002	0.002	0.002

	(0.011)	(0.011)	(0.011)	(0.012)
Father's Education-at Least Primary	-0.200**	-0.190*	-0.177*	-0.193*
	(0.095)	(0.097)	(0.099)	(0.101)
Mother's Education-at Least Primary	0.140	0.144	0.140	0.128
	(0.148)	(0.150)	(0.153)	(0.158)
Distance to the Nearest Primary School	0.041**	0.041**	0.045**	0.045**
	(0.021)	(0.021)	(0.021)	(0.021)
Household Size	0.010	0.006	0.006	0.009
	(0.017)	(0.018)	(0.018)	(0.019)
Number of Siblings	-0.030	-0.031	-0.030	-0.036
	(0.023)	(0.024)	(0.024)	(0.025)
Age of Father	0.004	0.004	0.004	0.004
	(0.004)	(0.004)	(0.004)	(0.004)
Age of Mother	0.002	0.003	0.003	0.003
	(0.006)	(0.006)	(0.006)	(0.006)
Sex of Household Head	0.025	0.087	0.077	0.103
	(0.099)	(0.103)	(0.107)	(0.112)
Resid. from Height-for-age eqn.	-0.152	-0.169	-0.188*	-0.170
	(0.105)	(0.104)	(0.108)	(0.117)
Year=2004	0.080	0.015	0.080	0.097
	(0.109)	(0.114)	(0.118)	(0.120)
Constant	-0.467	-0.355	-0.349	-0.388
	(0.389)	(0.400)	(0.408)	(0.422)
Athrho	-0.521***	-0.518***	-0.510***	-0.522***
	(0.053)	(0.054)	(0.055)	(0.057)
Observations	2241	2145	2053	1958

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: Site dummies were included in all equations to control for community fixed effects. Equation (I) was estimated using the entire sample. Equation (II) includes in the estimation sample only those who had at least 6 months of non-missing monthly rainfall records including the main rainy (agricultural) season during at least 1 of their critical years of development. The 5 and 6 year olds from Geblen and Haresaw and 1 year olds from Adele Keke did not fulfill these criteria and were excluded from the estimation sample for (II). Equation (III) includes in the estimation sample only those who had at least 6 months of non-missing monthly rainfall records including the main rainy (agricultural) season during at least 2 of their critical years of development. The 4, 5 and 6 year olds from Geblen and Haresaw, 1 and 2 year olds from Adele Keke and 4 year olds from Shumsha did not fulfill these criteria and were excluded from the estimation sample for (III). Equation (IV) includes in the estimation sample only those who had at least 6 months of non-missing monthly rainfall records including the main rainy (agricultural) season for all of the 3 critical years of development. The 3, 4, 5 and 6 year olds from Geblen and Haresaw, 1, 2 and 3 year olds from Adele Keke, 3 and 4 year olds from Shumsha, and 1 and 2 year olds from Imdibir did not fulfill these criteria and were excluded from the estimation sample for (IV).

Table B4. First-Stage Results for the Younger Cohort-Successively Adjusting the Estimation Sample for the Major Missing Rainfall Records

VARIABLES	(I)	(II)	(III)	(IV)
Substantial rainfall deficit at 1st year	0.047	0.021	0.052	0.059
	(0.223)	(0.226)	(0.233)	(0.244)
Substantial rainfall deficit at 2nd year	-0.401**	-0.396**	-0.379**	-0.371**
	(0.179)	(0.179)	(0.181)	(0.178)
Substantial rainfall deficit at 3rd year	-0.243	-0.240	-0.212	-0.246
	(0.178)	(0.178)	(0.180)	(0.181)
Substantial rainfall surplus at 1st year	-0.030	-0.088	-0.067	-0.046
	(0.178)	(0.196)	(0.197)	(0.198)
Substantial rainfall surplus at 2nd year	-0.227	-0.247	-0.224	-0.179
	(0.180)	(0.181)	(0.182)	(0.186)
Substantial rainfall surplus at 3rd year	-0.094	-0.090	-0.041	-0.147
	(0.251)	(0.251)	(0.261)	(0.273)
Height of Mother	0.028***	0.037***	0.037***	0.035***
	(0.010)	(0.011)	(0.011)	(0.011)
Height of Father	0.027***	0.026***	0.024***	0.026***
	(0.008)	(0.008)	(0.008)	(0.009)
Sex	-0.223***	-0.207***	-0.199	-0.176
	(0.119)	(0.121)	(0.124)	(0.128)
Age	-0.020	-0.035	-0.050	-0.054
	(0.032)	(0.034)	(0.034)	(0.035)
Household Size	0.016	0.025	0.021	0.015
	(0.025)	(0.024)	(0.024)	(0.025)
Number of Siblings	-0.034	-0.038	-0.032	-0.029
	(0.035)	(0.035)	(0.037)	(0.036)
Sex of Household Head	-0.102	-0.125	-0.173	-0.207
	(0.157)	(0.164)	(0.170)	(0.178)
Father's Education-at Least Primary	0.200	0.180	0.225	0.199
	(0.188)	(0.191)	(0.193)	(0.194)
Mother's Education-at Least Primary	-0.343	-0.326	-0.399	-0.404
	(0.260)	(0.263)	(0.253)	(0.259)
Age of Father	0.004	0.002	0.002	0.001
	(0.008)	(0.008)	(0.008)	(0.008)
Age of Mother	0.009	0.012	0.016	0.018
	(0.011)	(0.011)	(0.011)	(0.011)
Total Agricultural Land Area Owned	0.003	-0.000	0.002	-0.005
	(0.032)	(0.032)	(0.032)	(0.032)
Total Tropical Livestock Units Owned	-0.011	-0.012	-0.011	-0.009
	(0.010)	(0.010)	(0.011)	(0.010)
Year=2004	0.108	0.189	0.236	0.231
	(0.167)	(0.175)	(0.177)	(0.180)
Constant	-11.453***	-12.668***	-12.541***	-12.362***
	(1.901)	(1.943)	(2.014)	(2.087)
Observations	2241	2145	2053	1958

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: Site dummies were included in all equations to control for community fixed effects. Equations (I), (II), (III) and (IV) are first-stage results for equations (I), (II), (III) and (IV) in table B3, respectively.

Table B5. Marginal effects (at mean value) of Child's Height-for-age z-scores on the Choice Probabilities of Various Child Activities from the Two-Stage Bivariate Probit Results in Table B2

Model	p(stud=1|x)	p(stud=1, work=0|x)	p(stud=1, work=1|x)	p(stud=0, work=1|x)	p(stud=0, work=0|x)	p(work=1|x)
(I)	0.100***	-0.013	0.113***	-0.064*	-0.036***	0.048
	(0.038)	(0.023)	(0.032)	(0.035)	(0.012)	(0.031)
(II)	0.103***	-0.014	0.117***	-0.065*	-0.038***	0.052*
	(0.038)	(0.023)	(0.032)	(0.035)	(0.011)	(0.030)
(III)	0.093***	-0.018	0.111***	-0.055	-0.038***	0.056*
	(0.040)	(0.023)	(0.034)	(0.037)	(0.013)	(0.031)
(IV)	0.090**	-0.016	0.106***	-0.055	-0.036***	0.051
	(0.042)	(0.025)	(0.036)	(0.038)	(0.013)	(0.034)

*** p<0.01, ** p<0.05, * p<0.1

Notes: The partial effects reported in this table were calculated at the mean value of the child's height-for-age z- scores and other regressors and the standard errors were calculated by the delta method. Rows (I), (II), (III) and (IV) come from the results reported under equations (I), (II), (III) and (IV) in table B3, respectively.

Appendix C. The GLLAMM Model for Multinomial Logit

GLLAMM stands for Generalized Linear Latent and Mixed Models that includes a wide class of models elaborated in detail in Rabe-Hesketh and Skondrel (2008) and Rabe-Hesketh, Skrondal, and Pickles, (2004a, 2004b, 2004c, 2003). The GLLAMM models allow the unobserved heterogeneities at multiple levels of observation to be accounted for like a child and family levels in our case. They also recognize the nested structure of the observations as in our level-1 units of observation or children being nested or clustered under our level-2 units of observation that are families or households. In a panel data setting level 1 units are occasions or panel waves, level 2 units are children and level 3 units are families or households.

The multinomial logit model in this framework consists of three parts: the linear predictor, assumptions about the distribution of the latent variables for unobserved heterogeneities, and multinomial logit link. Assuming that the indirect utility under equation 1.21 in the main text is linear in its arguments, the linear predictor for a three level random intercept model for child i and household f at measurement occasion or panel wave w may be stated as,

$$v_{ifw}^{(j)} = \beta^{(j)'} x_{iw} + \delta^{(j)'} z_{fw} + \gamma^{(j)'} x_i + \lambda^{(j)'} z_f + \alpha_i + \eta_f \qquad \text{(C1)}$$

where j = 0,1,2,3 denotes categories of child activity choices, x_{iw} is a vector of child specific covariates that remain the same across alternatives but vary across children and over time, z_{fw} is a vector of family specific covariates that vary over time , x_i are time invariant child specific characteristics, z_f are time invariant family specific characteristics, α_i are child specific time invariant unobserved heterogeneities, η_f are time invariant family specific unobserved characteristics, and $\beta^{(j)}$, $\delta^{(j)}$, $\gamma^{(j)}$, and $\lambda^{(j)}$ are parameters. We do not have category specific covariates in our model.

Conditional on the random effects, the child activity choice, y_{ifw} , is assumed to have a multinomial logit distribution,

$$P(y_{ifw} = j \mid x_{iw}, z_{fw}, x_i, z_{f,} \alpha_i, \eta_f) = \frac{\exp\left\{v_{ifw}^{(j)}\right\}}{1 + \sum_{j=1}^{3} \exp\left\{v_{ifw}^{(j)}\right\}} \qquad \text{(C2)}$$

where, we have assumed all the random intercepts and the coefficients for the base category to be equal to zero for identification.

The random intercepts are assumed to be independently normally distributed.

$$\alpha_i^{'} = (\alpha_i^1, ..., \alpha_i^3)^{'} \underset{\longleftrightarrow}{\underline{iid}} \; N(0, \Sigma_\alpha); \qquad \text{(C3)}$$

$$\eta_i^{'} = (\eta_f^1, ..., \alpha_f^3)^{'} \underset{\longleftrightarrow}{\underline{iid}} \; N(0, \Sigma_\eta). \qquad \text{(C4)}$$

Since the choice probabilities stated under 1.C2 are conditional on the time-invariant unobserved child and family level heterogeneities, estimation of the model requires integrating over the distribution of the unobserved heterogeneities. Thus, the likelihood for the multinomial logit with child and family level random intercepts has the following form,

$$L(\omega) = \prod_{j=1}^{3} \int_\alpha \prod_{n=1}^{N} \prod_{w=1}^{2} \left\{ \int_\eta P(y_{ifw} = j \mid x_{iw}, z_{iw}, x_i, z_f, \alpha_i, \eta_f) f(\eta) d\eta \right\} f(\alpha) d\alpha \qquad \text{(C5)}$$

where, both f(η) and g(α) are standard normal densities and hence the random effects have a bivariate normal density and ω is a vector of parameters,

$$\omega' = (\beta^{(1)},...,\beta^{(3)},\delta^{(1)},...,\delta^{(3)},\lambda^{(1)},...,\lambda^{(3)},\gamma^{(1)},...,\gamma^{(3)},\Sigma_{\alpha},\Sigma_{\eta}).$$

Since the integrands in the likelihood function in C5 do not have closed from solutions, numerical approximations are needed to maximize it. The gllamm program in stata has the option of approximating the integrals by either Gauss-Hermite quadrature or adaptive quadrature. Adaptive quadrature is a method that extends Gauss Hermite quadrature by exploiting the posterior density of the unobserved heterogeneities which is approximately multivariate normal for large cluster sizes (Rabe-Hesketh, Skrondal, and Pickles, 2004c). The results reported under the last two columns of Table A7 and Table A8 were estimated using adaptive quadrature. This is implemented in stata by suitably arranging the data, defining level identifiers and feeding them to the GLLAMM program. To ensure the program recognizes the correct hierarchy of observations the identifiers are stated sequentially by starting with the lowest level.

Appendix D. Additional Summary Statistics and Results for Essay 2 from Ethiopian Data

Table D1. Summary Statistics for the Variables Used in the Models

	School Attendance				Weight-for-Height			
	DHS2000 (N=12502)		DHS2005 (N=11317)		DHS2000 (N=6427)		DHS2005 (N=3077)	
Variables	M	Std. Dev.	M	Std. Dev	M	Std. Dev.	M	Std. Dev
Child is attending school	0.35	0.48	0.39	0.49				
Weight-for-Height Z-scores					-1.81	1.24	-1.53	1.37
Percentage of wasted children					0.46	0.50	0.39	0.49
Mother's Desired Female Ratio	0.47	0.16	0.47	0.14	0.47	0.17	0.47	0.15
Actual Female Ratio	0.49	0.23	0.48	0.23	0.49	0.27	0.49	0.26
Gender Ratio Gap	0.02	0.21	0.01	0.21	0.02	0.25	0.02	0.23
Mother's Bargaining Power	0.00	0.89	0.00	0.89	0.00	0.87	0.00	0.90
Child is female	0.48	0.50	0.48	0.50	0.50	0.50	0.49	0.50
Age of Child	10.89	3.58	9.45	2.55	2.11	1.41	2.16	1.41
Household Size	7.31	2.07	7.16	1.93	6.40	2.02	6.47	1.96
Child's Birth-Order	3.90	2.35	3.75	2.34	4.64	2.55	4.62	2.49
First Wealth Quintile	0.12	0.33	0.29	0.45	0.13	0.34	0.27	0.44
Second Wealth Quintile	0.16	0.37	0.18	0.38	0.17	0.37	0.19	0.39
Fourth Wealth Quintile	0.20	0.40	0.17	0.38	0.21	0.41	0.18	0.39
Fifth Wealth Quintile	0.27	0.44	0.20	0.40	0.26	0.44	0.18	0.38
Mother's Body Mass Index					20.17	2.50	20.38	2.61
Gender Ratio of Mother's Siblings	0.50	0.25	0.47	0.26	0.48	0.24	0.47	0.25
Female Twin-Pair in the Family	0.01	0.09	0.01	0.09	0.01	0.11	0.00	0.07
Male Twin-Pair in the Family	0.01	0.10	0.01	0.10	0.01	0.09	0.00	0.07
Mixed Twin-Pair in the Family	0.01	0.12	0.01	0.11	0.01	0.09	0.02	0.12
Independence in Seeking External Relations	0.41	0.49	0.32	0.47	0.40	0.49	0.34	0.47
Independence in Child-care Decisions	0.35	0.48	0.32	0.47	0.33	0.47	0.34	0.47
Self-confidence in Talking to Husband	0.39	0.49	0.38	0.49	0.36	0.48	0.37	0.48
Independence in Sexual Decisions	0.45	0.50	0.49	0.50	0.43	0.50	0.50	0.50
Independence in Domestic Activities	0.38	0.49	0.40	0.49	0.34	0.47	0.39	0.49
Husband has more than 1 wife	0.84	0.37	0.85	0.36	0.87	0.33	0.87	0.34
Mother Listens to Radio	0.46	0.87	0.55	0.92	0.43	0.82	0.53	0.90
Mother is More Educated than Husband	0.05	0.22	0.07	0.25	0.06	0.24	0.08	0.27
Mother's age-Husband's age	-9.27	6.90	-8.78	6.65	-8.69	6.56	-8.27	6.84
Mother is Non-Muslim	0.60	0.49	0.58	0.49	0.62	0.49	0.60	0.49

Source: The Demographic and Health Surveys of Ethiopia Conducted in the Years 2000 and 2005.
Notes: M stands for mean.

Table D2. Factor Loadings (pattern matrix) and Unique Variances

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Uniqueness
Independence in Seeking Ext. Relations	0.666	-0.042	0.024	0.121	0.007	0.540
Independence in Child-care Decisions	0.690	-0.101	0.059	0.089	0.013	0.502
Self-confidence in Talking to Husband	0.689	-0.050	0.006	-0.008	-0.003	0.523
Independence in Sexual Decisions	0.590	0.091	-0.010	-0.101	-0.025	0.632
Independence in Domestic Activities	0.638	-0.052	-0.031	-0.127	0.001	0.573
Mother Listens to Radio	0.202	0.263	-0.172	-0.005	0.010	0.861
Mother is More Educated than Husband	0.082	0.172	-0.228	0.032	0.013	0.911
Mother is Non-Muslim	0.070	0.283	0.048	0.102	-0.028	0.901
Mother's age-Husband's age	0.051	0.206	0.147	-0.066	0.030	0.928
Husband has Just 1 Wife	0.049	0.297	0.158	-0.011	0.003	0.884

Observations= 40,178.

Table D3. Factor Analysis/Eigen Values

Factor	Eigen value	Difference	Proportion	Cumulative
Factor1	2.207	1.872	1.090	1.090
Factor2	0.335	0.199	0.166	1.256
Factor3	0.136	0.071	0.067	1.323
Factor4	0.065	0.062	0.032	1.355
Factor5	0.003	0.100	0.001	1.356
Factor6	-0.097	0.018	-0.048	1.308
Factor7	-0.115	0.017	-0.057	1.251
Factor8	-0.132	0.034	-0.065	1.186
Factor9	-0.166	0.046	-0.082	1.104
Factor10	-0.211	.	-0.104	1.000

Notes: Some of the eigenvalues are negative and hence the cumulative proportion is greater than 1 because in principal factor analysis that was followed here, the sample covariance matrix is not guaranteed to be positive semidefinte as opposed to the principal component analysis (Rencher, 1995).

Table D4. First Stage Results for the Models of Schooling- OLS Regression of Gender Ratio Gap on Instruments and Other Covariates

	For Baseline Results		For Main Results	
VARIABLES	Siblings	Twins	Siblings	Twins
Gender Ratio of Mother's Siblings	0.037***	0.075	0.037***	0.070
	(0.008)	(0.077)	(0.008)	(0.077)
Female Twin-Pair in the Family	0.087***	0.157***	0.087***	0.164***
	(0.023)	(0.042)	(0.023)	(0.043)
Male Twin-Pair in the Family	-0.102***	-0.141***	-0.102***	-0.137***
	(0.024)	(0.043)	(0.024)	(0.042)
Child is female	0.147***	0.025	0.147***	0.025
	(0.003)	(0.016)	(0.003)	(0.016)
Mother's Bargaining Power			0.000	-0.020
			(0.002)	(0.019)
Age of Child	-0.000	-0.002	-0.000	-0.002
	(0.000)	(0.006)	(0.000)	(0.006)
Household Size	-0.001	0.003	-0.001	0.002
	(0.001)	(0.009)	(0.001)	(0.009)
Child's Birth-order	-0.000	-0.000	-0.000	-0.001
	(0.001)	(0.007)	(0.001)	(0.006)
First Wealth Quintile	0.007	-0.013	0.007	-0.008
	(0.006)	(0.073)	(0.006)	(0.072)
Second Wealth Quintile	0.010	0.040	0.010	0.039
	(0.007)	(0.078)	(0.007)	(0.079)
Fourth Wealth Quintile	-0.001	-0.017	-0.001	0.000
	(0.007)	(0.084)	(0.007)	(0.085)
Fifth Wealth Quintile	-0.000	-0.023	-0.000	-0.012
	(0.006)	(0.071)	(0.006)	(0.071)
Survey Round is 2005	-0.006	0.023	-0.006	0.024
	(0.004)	(0.040)	(0.004)	(0.039)
Constant	-0.064***	-0.047	-0.064***	-0.050
	(0.011)	(0.147)	(0.011)	(0.145)
F-stat for joint sig. of instruments	15.93***	11.61***	15.93***	11.37***
	(p=0.00)	(p=0.00)	(p=0.00)	(p=0.00)
Observations	23819	252	23819	252

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table D5. First Stage Results Weight-For-Height Models- OLS Regression of Gender Ratio Gap on Instruments and Other Covariates

	For Baseline Results		For Main Results	
VARIABLES	Siblings	Twins	Siblings	Twins
Gender Ratio of Mother's Siblings	0.018	0.094	0.018	0.091
	(0.012)	(0.112)	(0.012)	(0.112)
Female Twin-Pair in the Family	0.054*	0.144***	0.053*	0.145***
	(0.032)	(0.050)	(0.032)	(0.050)
Male Twin-Pair in the Family	-0.083**	-0.117*	-0.084**	-0.116*
	(0.035)	(0.067)	(0.035)	(0.068)

Child is female	0.209***	-0.000	0.209***	-0.001
	(0.005)	(0.006)	(0.005)	(0.007)
Mother's Bargaining Power			-0.005	-0.012
			(0.003)	(0.027)
Age of Child	0.000	-0.014	0.000	-0.013
	(0.001)	(0.019)	(0.001)	(0.018)
Household Size	-0.001	-0.010	-0.001	-0.011
	(0.002)	(0.010)	(0.002)	(0.011)
Child's Birth-order	0.001	-0.011*	0.001	-0.010
	(0.001)	(0.006)	(0.001)	(0.006)
First Wealth Quintile	0.010	-0.176**	0.009	-0.178**
	(0.009)	(0.070)	(0.009)	(0.071)
Second Wealth Quintile	0.006	-0.053	0.006	-0.052
	(0.009)	(0.070)	(0.009)	(0.070)
Fourth Wealth Quintile	-0.005	-0.141	-0.005	-0.139
	(0.009)	(0.099)	(0.009)	(0.100)
Fifth Wealth Quintile	-0.013	-0.177***	-0.012	-0.166**
	(0.009)	(0.063)	(0.009)	(0.066)
Mother's Body Mass Index	0.001	0.009	0.001	0.009
	(0.001)	(0.013)	(0.001)	(0.013)
Survey Round is 2005	-0.008	-0.007	-0.007	0.002
	(0.006)	(0.045)	(0.006)	(0.053)
Constant	-0.110***	0.041	-0.112***	0.035
	(0.024)	(0.288)	(0.024)	(0.287)
F-stat for joint sig. of instruments	3.66**	6.71***	3.65**	6.56***
	(0.01)	(0.00)	(0.01)	(0.00)
Observations	9504	148	9504	148

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table D6. Linear Probability Models for School Attendance-Baseline Results from Siblings' Data

	(1)	(2)	(3)	(4)
VARIABLES	OLS	GLS	2SLS	IV-GLS
Child is female	-0.057***	-0.062***	-0.168***	-0.105***
	(0.006)	(0.006)	(0.030)	(0.019)
Gender Ratio Gap	0.029	0.051**	1.053***	0.977***
	(0.023)	(0.022)	(0.282)	(0.300)
Gender Ratio Gap*Female	0.000	-0.033	-0.451	-0.621*
	(0.031)	(0.029)	(0.356)	(0.352)
Age of Child	0.044***	0.043***	0.043***	0.042***
	(0.001)	(0.001)	(0.001)	(0.001)
Household Size	-0.002	-0.002	-0.004*	-0.004
	(0.002)	(0.002)	(0.002)	(0.003)
Child's Birth-order	-0.007***	-0.007***	-0.009***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
First Wealth Quintile	-0.146***	-0.142***	-0.150***	-0.145***
	(0.011)	(0.011)	(0.010)	(0.013)
Second Wealth Quintile	-0.041***	-0.040***	-0.048***	-0.044***

	(0.012)	(0.011)	(0.010)	(0.013)
Fourth Wealth Quintile	0.042***	0.041***	0.043***	0.041***
	(0.012)	(0.011)	(0.010)	(0.012)
Fifth Wealth Quintile	0.230***	0.230***	0.233***	0.238***
	(0.011)	(0.011)	(0.010)	(0.012)
Survey Round is 2005	0.145***	0.145***	0.151***	0.149***
	(0.007)	(0.007)	(0.007)	(0.008)
Constant	-0.097***	-0.097***	-0.020	-0.035
	(0.018)	(0.017)	(0.036)	(0.037)
Observations	23819	23819	23819	23819
Number of Mother_id		9911		9911

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note: The sample consists of children from age 6 to 18 in the combined data from the two rounds of DHS-Ethiopia. Only those children with married mother with at least two children are included.

Table D7. Linear Probability Models for School Attendance-Baseline Results from Twins' Data

	(1)	(2)	(3)	(4)
VARIABLES	OLS	GLS	2SLS	IV-GLS
Child is female	0.047	-0.023	-0.109	-0.129
	(0.062)	(0.051)	(0.094)	(0.094)
Gender Ratio Gap	-0.250	-0.060	-0.128	0.010
	(0.205)	(0.186)	(0.515)	(0.518)
Gender Ratio Gap*Female	0.285	-0.013	1.482*	1.354*
	(0.251)	(0.146)	(0.753)	(0.756)
Age of Child	0.049***	0.045***	0.047***	0.048***
	(0.011)	(0.012)	(0.009)	(0.009)
Household Size	-0.023	-0.024	-0.017	-0.020
	(0.018)	(0.018)	(0.016)	(0.016)
Child's Birth-Order	-0.009	-0.025*	-0.003	-0.005
	(0.014)	(0.014)	(0.012)	(0.013)
First Wealth Quintile	-0.128	-0.136	-0.103	-0.110
	(0.125)	(0.136)	(0.105)	(0.106)
Second Wealth Quintile	0.012	-0.028	0.007	-0.004
	(0.114)	(0.123)	(0.113)	(0.114)
Fourth Wealth Quintile	0.168	0.166	0.193*	0.222**
	(0.142)	(0.154)	(0.107)	(0.108)
Fifth Wealth Quintile	0.313***	0.295**	0.354***	0.348***
	(0.109)	(0.120)	(0.089)	(0.089)
Survey Round is 2005	0.164**	0.140*	0.154**	0.158**
	(0.081)	(0.085)	(0.070)	(0.071)
Constant	-0.124	0.077	-0.188	-0.144
	(0.182)	(0.206)	(0.212)	(0.213)
Observations	252	252	252	252
Number of Mother_id		121		121

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The sample consists of living twin-pairs from age 6 to 18 in the combined data from the two rounds of DHS-Ethiopia. Only those children with married mother with at least two children are included.

Table D8. Models for Child's Weight-for-Height Z-Scores –Baseline Results from Siblings' Data

VARIABLES	(1) OLS	(2) GLS	(3) 2SLS	(4) IV-GLS
Child is female	-0.011	-0.006	0.360	0.167
	(0.028)	(0.027)	(0.393)	(0.285)
Gender Ratio Gap	0.014	0.039	1.582	1.402
	(0.087)	(0.085)	(1.172)	(1.158)
Gender Ratio Gap*Female	-0.059	-0.097	-5.641*	-3.921
	(0.120)	(0.118)	(3.014)	(2.586)
Age of Child	-0.180***	-0.172***	-0.191***	-0.180***
	(0.009)	(0.009)	(0.012)	(0.011)
Household Size	0.036***	0.037***	0.016	0.023
	(0.009)	(0.009)	(0.015)	(0.014)
Child's Birth-Order	-0.050***	-0.050***	-0.089***	-0.077***
	(0.007)	(0.007)	(0.022)	(0.019)
First Wealth Quintile	0.019	0.028	0.002	0.016
	(0.045)	(0.045)	(0.048)	(0.049)
Second Wealth Quintile	-0.044	-0.036	-0.018	-0.018
	(0.043)	(0.043)	(0.050)	(0.050)
Fourth Wealth Quintile	0.053	0.055	0.059	0.060
	(0.041)	(0.041)	(0.045)	(0.046)
Fifth Wealth Quintile	0.262***	0.272***	0.274***	0.290***
	(0.040)	(0.040)	(0.045)	(0.046)
Mother's Body Mass Index	0.093***	0.093***	0.099***	0.097***
	(0.006)	(0.006)	(0.007)	(0.007)
Survey Round is 2005	0.283***	0.280***	0.246***	0.260***
	(0.030)	(0.030)	(0.040)	(0.038)
Constant	-3.370***	-3.404***	-3.025***	-3.127***
	(0.126)	(0.126)	(0.215)	(0.213)
Observations	9504	9504	9504	9504
Number of Mother_id		6425		6425

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The sample consists of children under age 5 in the combined data from the two rounds of DHS-Ethiopia. Only those children with married mother with at least two children are included.

Table D9. Models for Child's Weight-for-Height Z-Scores –Baseline Results from Twins' Data

VARIABLES	(1) OLS	(2) GLS	(3) 2SLS	(4) IV-GLS
Child is female	-0.105	0.089	0.206	0.194
	(0.215)	(0.167)	(0.319)	(0.167)
Gender Ratio Gap	1.880**	0.447	0.828	-0.164
	(0.823)	(0.710)	(2.128)	(2.347)
Gender Ratio Gap*Female	-2.982**	-0.968	-4.523*	-3.014
	(1.215)	(1.063)	(2.348)	(2.883)
Age of Child	0.199**	0.199**	0.186**	0.185*
	(0.084)	(0.087)	(0.072)	(0.098)
Household Size	0.013	0.054	-0.009	0.015
	(0.098)	(0.104)	(0.064)	(0.096)
Child's Birth-Order	-0.079	-0.104*	-0.119**	-0.133**
	(0.049)	(0.054)	(0.051)	(0.056)
First Wealth Quintile	-0.238	-0.315	-0.585	-0.626
	(0.539)	(0.542)	(0.469)	(0.611)
Second Wealth Quintile	-0.517	-0.477	-0.689*	-0.637
	(0.411)	(0.457)	(0.358)	(0.493)
Fourth Wealth Quintile	1.336**	1.159**	1.155***	1.081*
	(0.541)	(0.547)	(0.439)	(0.580)
Fifth Wealth Quintile	1.238***	1.096***	0.929**	0.845*
	(0.402)	(0.414)	(0.402)	(0.512)
Mother's Body Mass Index	0.042	0.060	0.040	0.050
	(0.049)	(0.052)	(0.047)	(0.066)
Survey Round is 2005	0.067	0.123	-0.011	0.013
	(0.347)	(0.360)	(0.259)	(0.367)
Constant	-3.414**	-4.088***	-2.796**	-3.131*
	(1.330)	(1.378)	(1.164)	(1.634)
Observations	148	148	148	148
Number of Mother_id		73		73

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The sample consists of living twin-pairs under age 5 in the combined data from the two rounds of DHS-Ethiopia. Only those children with married mother with at least two children are included.

Table D10: Gender Preferences by Husband and Wife

	Percent of Cases			
	2000 (Obs=2168)		2005 (Obs=4752)	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Wife Prefers More Boys than Husband	33.90	47.35	32.49	46.84
Wife Prefers Less Boys than Husband	43.17	49.54	43.77	49.62
Husband and Wife Prefer Same Number of Boys	22.92	42.04	23.74	42.55
Wife Prefers More Girls than Husband	41.10	49.21	37.50	48.42
Wife Prefers Less Girls than Husband	32.29	46.77	34.62	47.58
Husband and Wife Prefer Same Number of Girls	26.61	44.20	27.88	44.85
Wife Prefers More Children than Husband	38.51	48.67	38.09	48.57
Wife Prefers Less Children than Husband	40.08	49.02	43.06	49.52
Husband and Wife Prefer Same Number of Children	21.40	41.02	18.86	39.12

Source: Demographic and Health Surveys of Ethiopia

Appendix E. Results for Essay 2 from Indian Demographic and Health Survey of 2005/06

Table E1. Summary Statistics for the Variables Used in the Models-Indian Data

	School Attendance (N=91899)		Weight-for-Height (N=9504)	
Variables	Mean	Std. Dev.	Mean	Std. Dev.
Child is attending school	0.75	0.43		
Weight-for-Height Z-scores			-1.83	1.19
Mother's Desired Female Ratio	0.44	0.17	0.44	0.17
Actual Female Ratio	0.48	0.26	0.51	0.29
Gender Ratio Gap	0.04	0.25	0.06	0.28
Mother's Bargaining Power (de-meaned)	0.00	0.91	0.00	0.87
Child is female	0.48	0.50	0.48	0.50
Age of Child	11.63	3.62	2.30	1.37
Household Size	6.15	2.07	5.79	1.88
Child's Birth-Order	2.84	1.75	3.40	1.94
First Wealth Quintile	0.16	0.37	0.24	0.43
Second Wealth Quintile	0.18	0.39	0.21	0.41
Fourth Wealth Quintile	0.23	0.42	0.20	0.40
Fifth Wealth Quintile	0.22	0.41	0.13	0.34
Mother's Body Mass Index			20.24	3.39
Gender Ratio of Mother's Siblings				
Female Twin-Pair in the Family	0.01	0.10	0.01	0.10
Male Twin-Pair in the Family	0.01	0.11	0.01	0.10
Mixed Twin-Pair in the Family	0.01	0.11	0.01	0.09
Independence in Seeking External Relations	0.63	0.48	0.66	0.47
Independence in Child-care Decisions	0.56	0.50	0.58	0.49
Self-confidence in Talking to Husband	0.61	0.49	0.65	0.48
Independence in Sexual Decisions	0.77	0.42	0.82	0.38
Independence in Domestic Activities	0.74	0.44	0.77	0.42
Husband has just 1 wife	0.98	0.14	0.98	0.15
Mother Listens to Radio	0.98	1.19	0.79	1.11
Mother is More Educated than Husband	0.13	0.34	0.17	0.37
Mother's age-Husband's age	-6.07	5.87	-6.01	6.17
Mother is Non-Muslim	0.98	0.14	0.99	0.10

Source: Demographic and Health Surveys of India Conducted in 2005/06.

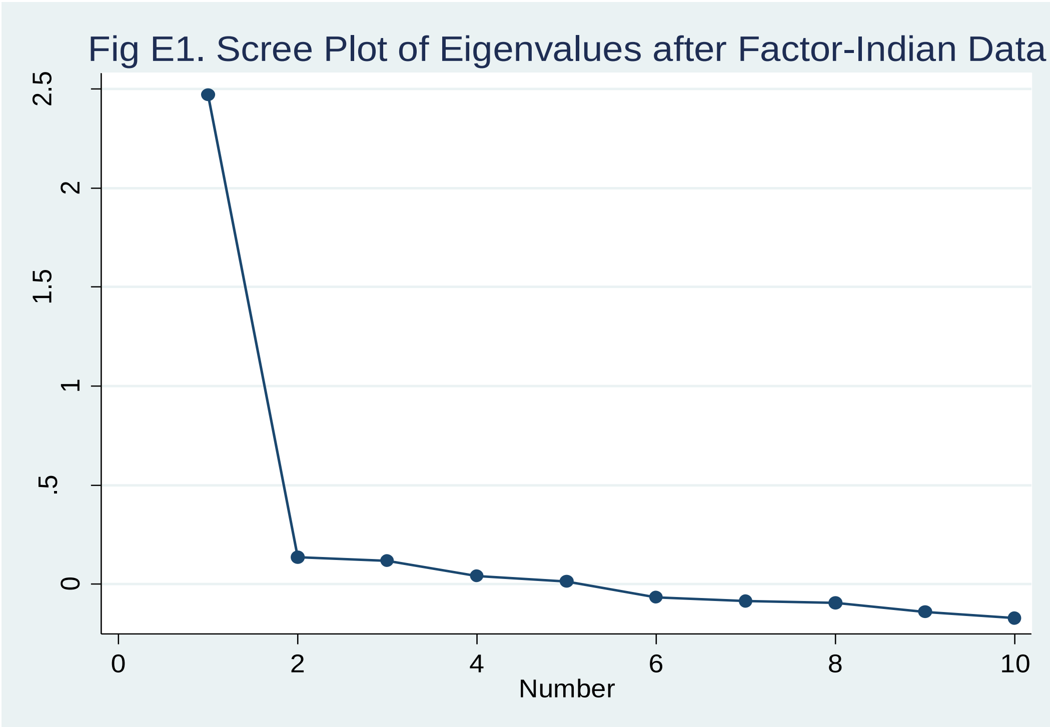


Table E2. First Stage Results for the Models of Schooling- OLS Regression of Gender Ratio Gap on Instruments and Other Covariates- Indian Data

	For Baseline Results		For Main Results	
VARIABLES	Siblings	Twins	Siblings	Twins
Female Twin-Pair in the Family	0.092***	0.140***	0.092***	0.137***
	(0.012)	(0.020)	(0.012)	(0.021)
Male Twin-Pair in the Family	-0.099***	-0.163***	-0.098***	-0.157***
	(0.012)	(0.020)	(0.012)	(0.020)
Child is female	0.202***	0.015**	0.201***	0.019***
	(0.002)	(0.006)	(0.002)	(0.007)
Mother's Bargaining Power			-0.007***	-0.005
			(0.001)	(0.009)
Age of Child	-0.000	-0.003	-0.000*	-0.003
	(0.000)	(0.002)	(0.000)	(0.002)
Household Size	0.006***	0.006	0.006***	0.006
	(0.001)	(0.004)	(0.001)	(0.004)
Child's Birth-order	0.010***	0.015***	0.010***	0.014***
	(0.001)	(0.004)	(0.001)	(0.004)
First Wealth Quintile	0.018***	0.054*	0.019***	0.054*
	(0.004)	(0.029)	(0.004)	(0.029)
Second Wealth Quintile	0.010***	0.020	0.010***	0.011
	(0.004)	(0.027)	(0.004)	(0.027)
Fourth Wealth Quintile	-0.010***	-0.000	-0.009**	0.002
	(0.003)	(0.023)	(0.003)	(0.024)
Fifth Wealth Quintile	-0.018***	0.023	-0.015***	0.026

	(0.003)	(0.026)	(0.004)	(0.026)
Constant	-0.121***	-0.031	-0.118***	-0.030
	(0.005)	(0.042)	(0.005)	(0.043)
Observations	91899	1377	91899	1377

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table E3. First Stage Results for Weight-for-Height Models- OLS Regression of Gender Ratio Gap on Instruments and Other Covariates- Indian Data

	For Baseline Results		For Main Results	
VARIABLES	Siblings	Twins	Siblings	Twins
Female Twin-Pair in the Family	0.101***	0.164***	0.104***	0.182***
	(0.023)	(0.044)	(0.023)	(0.049)
Male Twin-Pair in the Family	-0.109***	-0.210***	-0.101***	-0.164***
	(0.019)	(0.044)	(0.020)	(0.053)
Child is female	0.256***	0.012	0.256***	0.031**
	(0.004)	(0.019)	(0.004)	(0.013)
Mother's Bargaining Power			0.000	0.004
			(0.002)	(0.024)
Age of Child	-0.004***	0.005	-0.004***	0.003
	(0.001)	(0.012)	(0.001)	(0.013)
Household Size	0.007***	0.012	0.007***	0.011
	(0.001)	(0.009)	(0.001)	(0.009)
Child's Birth-order	0.006***	-0.007	0.006***	-0.005
	(0.001)	(0.009)	(0.001)	(0.010)
First Wealth Quintile	0.019***	0.070	0.019***	0.067
	(0.006)	(0.058)	(0.006)	(0.057)
Second Wealth Quintile	0.018***	0.034	0.018***	0.049
	(0.006)	(0.057)	(0.006)	(0.059)
Fourth Wealth Quintile	0.000	-0.027	0.000	-0.024
	(0.006)	(0.054)	(0.006)	(0.054)
Fifth Wealth Quintile	-0.015**	-0.075	-0.015*	-0.071
	(0.007)	(0.056)	(0.007)	(0.055)
Mother's Body Mass Index	-0.001	-0.002	-0.001	-0.003
	(0.001)	(0.005)	(0.001)	(0.005)
Constant	-0.104***	0.075	-0.104***	0.066
	(0.015)	(0.130)	(0.015)	(0.126)
Observations	20975	317	20975	317

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table E4. Linear Probability Models for School Attendance-Baseline Results from Siblings' Data

VARIABLES	(1) OLS	(2) GLS	(3) 2SLS	(4) IV-GLS
Child is female	-0.029***	-0.032***	-0.021	-0.007
	(0.003)	(0.003)	(0.028)	(0.024)
Gender Ratio Gap	0.022***	0.029***	0.474***	0.508***
	(0.009)	(0.009)	(0.061)	(0.058)
Gender Ratio Gap*Female	-0.036***	-0.049***	-0.725***	-0.853***
	(0.013)	(0.012)	(0.155)	(0.175)
Age of Child	-0.018***	-0.019***	-0.019***	-0.020***
	(0.000)	(0.000)	(0.000)	(0.000)
Household Size	-0.017***	-0.017***	-0.020***	-0.020***
	(0.001)	(0.001)	(0.001)	(0.001)
Child's Birth-order	-0.013***	-0.009***	-0.017***	-0.014***
	(0.001)	(0.001)	(0.001)	(0.001)
First Wealth Quintile	-0.160***	-0.162***	-0.163***	-0.164***
	(0.006)	(0.006)	(0.005)	(0.006)
Second Wealth Quintile	-0.055***	-0.056***	-0.056***	-0.057***
	(0.006)	(0.006)	(0.005)	(0.005)
Fourth Wealth Quintile	0.065***	0.065***	0.067***	0.068***
	(0.005)	(0.005)	(0.004)	(0.005)
Fifth Wealth Quintile	0.153***	0.156***	0.155***	0.156***
	(0.005)	(0.004)	(0.005)	(0.005)
Constant	1.107***	1.105***	1.169***	1.172***
	(0.009)	(0.008)	(0.010)	(0.010)
Observations	91899	91899	91899	91899
Number of Mother_id		40486		40486

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The sample consists of children from age 6 to 18 in the DHS-India. Only those children with married mother with at least two children are included.

Table E5. Linear Probability Models for School Attendance-Baseline Results from Twins' Data

VARIABLES	(1) OLS	(2) GLS	(3) 2SLS	(4) IV-GLS
Child is female	-0.039	-0.033	0.006	-0.003
	(0.030)	(0.028)	(0.049)	(0.046)
Gender Ratio Gap	-0.021	-0.025	0.084	-0.158
	(0.079)	(0.079)	(0.185)	(4.644)
Gender Ratio Gap*Female	0.026	0.023	-0.431	-0.225
	(0.120)	(0.121)	(0.300)	(0.703)
Age of Child	-0.007	-0.008	-0.008**	-0.080***
	(0.005)	(0.005)	(0.004)	(0.024)
Household Size	-0.003	-0.002	-0.002	0.011
	(0.006)	(0.007)	(0.006)	(0.252)
Child's Birth-Order	-0.031***	-0.032***	-0.032***	-0.047***
	(0.008)	(0.009)	(0.008)	(0.014)

First Wealth Quintile	-0.117**	-0.117**	-0.114**	-0.162
	(0.055)	(0.055)	(0.047)	(2.096)
Second Wealth Quintile	-0.019	-0.016	-0.017	-0.030
	(0.052)	(0.052)	(0.042)	(1.921)
Fourth Wealth Quintile	0.055	0.063	0.052	0.121
	(0.044)	(0.045)	(0.038)	(1.745)
Fifth Wealth Quintile	0.144***	0.147***	0.154***	0.249
	(0.047)	(0.047)	(0.041)	(1.793)
Constant	0.829***	0.830***	0.848***	1.597
	(0.080)	(0.080)	(0.069)	(2.193)
Observations	1377	1377	1377	1377
Number of Mother_id		670		670

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The sample consists of living twin-pairs (and triplets) from age 6 to 18 in the DHS-India. Only those children with married mother with at least two children are included.

Table E6. Models for Child's Weight-for-Height Z-Scores –Baseline Results from Siblings' Data

VARIABLES	(1) OLS	(2) GLS	(3) 2SLS	(4) IV-GLS
Child is female	-0.030*	-0.030*	-0.000	-0.019
	(0.018)	(0.017)	(0.223)	(0.193)
Gender Ratio Gap	-0.040	-0.021	0.646**	0.658**
	(0.046)	(0.045)	(0.274)	(0.297)
Gender Ratio Gap*Female	-0.088	-0.106	-1.132	-0.995
	(0.067)	(0.065)	(1.098)	(1.043)
Age of Child	-0.119***	-0.116***	-0.120***	-0.117***
	(0.006)	(0.006)	(0.007)	(0.007)
Household Size	0.011**	0.011**	0.007	0.007
	(0.006)	(0.006)	(0.006)	(0.006)
Child's Birth-Order	-0.049***	-0.052***	-0.058***	-0.060***
	(0.006)	(0.006)	(0.008)	(0.008)
First Wealth Quintile	-0.282***	-0.281***	-0.279***	-0.280***
	(0.025)	(0.025)	(0.029)	(0.031)
Second Wealth Quintile	-0.140***	-0.142***	-0.137***	-0.141***
	(0.026)	(0.025)	(0.029)	(0.031)
Fourth Wealth Quintile	0.104***	0.105***	0.106***	0.107***
	(0.026)	(0.025)	(0.024)	(0.026)
Fifth Wealth Quintile	0.327***	0.327***	0.325***	0.326***
	(0.031)	(0.031)	(0.031)	(0.034)
Mother's Body Mass Index	0.068***	0.068***	0.068***	0.068***
	(0.003)	(0.003)	(0.003)	(0.003)
Constant	-2.773***	-2.775***	-2.682***	-2.690***
	(0.065)	(0.064)	(0.069)	(0.074)
Observations	20975	20975	20975	20975
Number of Mother_id		14909		14909

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The sample consists of children under age 5 in the DHS-India. Only those children with married mother with at least two children are included.

Table E7. Models for Child's Weight-for-Height Z-Scores –Baseline Results from Twins' Data

	(1)	(2)	(3)	(4)
VARIABLES	OLS	GLS	2SLS	IV-GLS
Child is female	0.067	0.131	0.576*	0.502***
	(0.171)	(0.171)	(0.293)	(0.177)
Gender Ratio Gap	-0.179	-0.023	-0.666	0.852
	(0.554)	(0.477)	(0.914)	(1.276)
Gender Ratio Gap*Female	-0.358	-0.408	-2.162	-3.772***
	(0.650)	(0.544)	(1.322)	(1.407)
Age of Child	-0.002	-0.004	-0.016	-0.043
	(0.069)	(0.071)	(0.055)	(0.139)
Household Size	-0.023	0.005	0.000	-0.003
	(0.055)	(0.055)	(0.046)	(0.096)
Child's Birth-Order	0.001	-0.036	-0.035	-0.046
	(0.043)	(0.039)	(0.046)	(0.048)
First Wealth Quintile	-0.146	-0.125	-0.079	-0.021
	(0.310)	(0.320)	(0.257)	(0.646)
Second Wealth Quintile	-0.158	-0.090	-0.045	-0.005
	(0.250)	(0.250)	(0.263)	(0.660)
Fourth Wealth Quintile	0.464*	0.424*	0.422*	0.319
	(0.239)	(0.242)	(0.240)	(0.605)
Fifth Wealth Quintile	0.775***	0.705***	0.607**	0.400
	(0.257)	(0.250)	(0.269)	(0.646)
Mother's Body Mass Index	0.007	0.009	0.005	0.069
	(0.029)	(0.029)	(0.022)	(0.058)
Constant	-2.179***	-2.275***	-2.142***	-3.226**
	(0.735)	(0.723)	(0.560)	(1.423)
Observations	317	317	317	317
Number of Mother_id		152		152

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: The sample consists of living twin-pairs under age 5 in the DHS-India. Only those children with married mother with at least two children are included.

Table E8. Linear Probability Models for School Attendance-Results from Siblings' Data

	(1)	(2)	(3)	(4)
VARIABLES	OLS	GLS	2SLS	IV-GLS
Child is female	-0.028***	-0.032***	-0.026	-0.011
	(0.003)	(0.003)	(0.029)	(0.025)
Gender Ratio Gap	0.023***	0.030***	0.475***	0.505***
	(0.009)	(0.009)	(0.062)	(0.059)
Mother's Bargaining Power	0.005**	0.005**	-0.000	-0.001

	(0.002)	(0.002)	(0.003)	(0.003)
Gender Ratio Gap*Female	-0.035***	-0.048***	-0.675***	-0.801***
	(0.013)	(0.012)	(0.161)	(0.182)
Gender Ratio Gap*Bargaining Power	-0.013	-0.013	-0.097**	-0.117***
	(0.010)	(0.010)	(0.042)	(0.045)
Gender Ratio Gap*Bargaining Power*Female	0.035**	0.039***	0.178***	0.200***
	(0.014)	(0.013)	(0.060)	(0.059)
Age of Child	-0.018***	-0.019***	-0.019***	-0.020***
	(0.000)	(0.000)	(0.000)	(0.000)
Household Size	-0.017***	-0.017***	-0.020***	-0.020***
	(0.001)	(0.001)	(0.001)	(0.001)
Child's Birth-order	-0.012***	-0.009***	-0.017***	-0.014***
	(0.001)	(0.001)	(0.001)	(0.001)
First Wealth Quintile	-0.161***	-0.162***	-0.164***	-0.165***
	(0.006)	(0.006)	(0.005)	(0.006)
Second Wealth Quintile	-0.055***	-0.056***	-0.056***	-0.057***
	(0.006)	(0.006)	(0.005)	(0.005)
Fourth Wealth Quintile	0.064***	0.064***	0.066***	0.067***
	(0.005)	(0.005)	(0.004)	(0.005)
Fifth Wealth Quintile	0.150***	0.153***	0.152***	0.153***
	(0.005)	(0.005)	(0.005)	(0.005)
Constant	1.105***	1.102***	1.165***	1.166***
	(0.009)	(0.008)	(0.010)	(0.010)
Observations	91899	91899	91899	91899
Number of Mother_id		40486		40486

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The sample consists of children from age 6 to 18 in the DHS-India. Only those children with married mother with at least two children are included.

Table E9. Linear Probability Models for School Attendance-Results from Twins' Data

VARIABLES	(1) OLS	(2) GLS	(3) 2SLS	(4) IV-GLS
Child is female	-0.033	-0.030	0.007	0.016
	(0.030)	(0.028)	(0.050)	(0.040)
Gender Ratio Gap	0.000	-0.005	0.249	0.259
	(0.079)	(0.079)	(0.190)	(0.249)
Mother's Bargaining Power	0.043**	0.041**	0.038**	0.037
	(0.018)	(0.018)	(0.019)	(0.028)
Gender Ratio Gap*Female	0.010	0.013	-0.609*	-0.683
	(0.121)	(0.122)	(0.311)	(0.442)
Gender Ratio Gap*Bargaining Power	-0.185**	-0.181**	-0.199	-0.165
	(0.094)	(0.090)	(0.201)	(0.282)
Gender Ratio Gap*Bargaining				

Power*Female	0.143	0.128	0.203	0.157
	(0.126)	(0.118)	(0.283)	(0.359)
Age of Child	-0.007	-0.007	-0.008**	-0.010*
	(0.005)	(0.005)	(0.004)	(0.006)
Household Size	-0.001	-0.001	-0.002	-0.001
	(0.006)	(0.006)	(0.006)	(0.009)
Child's Birth-Order	-0.029***	-0.030***	-0.032***	-0.032***
	(0.008)	(0.009)	(0.008)	(0.009)
First Wealth Quintile	-0.121**	-0.120**	-0.123***	-0.123*
	(0.054)	(0.055)	(0.047)	(0.072)
Second Wealth Quintile	-0.015	-0.012	-0.015	-0.010
	(0.051)	(0.051)	(0.043)	(0.066)
Fourth Wealth Quintile	0.054	0.060	0.050	0.057
	(0.044)	(0.045)	(0.038)	(0.060)
Fifth Wealth Quintile	0.128***	0.132***	0.137***	0.144**
	(0.047)	(0.047)	(0.041)	(0.063)
Constant	0.806***	0.809***	0.847***	0.867***
	(0.079)	(0.080)	(0.069)	(0.100)
Observations	1377	1377	1377	1377
Number of Mother_id		670		670

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The sample consists of living twin-pairs (and triplets) from age 6 to 18 in the DHS-India. Only those children with married mother with at least two children are included.

Table E10. Models for Child's Weight-for-Height Z-Scores -Results from Siblings' Data

	(1)	(2)	(3)	(4)
VARIABLES	OLS	GLS	2SLS	IV-GLS
Child is female	-0.031*	-0.031*	-0.000	0.005
	(0.018)	(0.017)	(0.230)	(0.196)
Gender Ratio Gap	-0.042	-0.023	0.676**	0.673**
	(0.046)	(0.045)	(0.278)	(0.299)
Mother's Bargaining Power	-0.037***	-0.035***	-0.022	-0.017
	(0.011)	(0.011)	(0.018)	(0.017)
Gender Ratio Gap*Female	-0.084	-0.103	-1.173	-1.158
	(0.067)	(0.065)	(1.119)	(1.052)
Gender Ratio Gap*Bargaining Power	0.067	0.052	-0.001	0.078
	(0.052)	(0.051)	(0.231)	(0.239)
Gender Ratio Gap*Bargaining Power*Female	-0.081	-0.059	-0.168	-0.248
	(0.072)	(0.070)	(0.320)	(0.306)
Age of Child	-0.119***	-0.116***	-0.120***	-0.117***
	(0.006)	(0.006)	(0.007)	(0.007)
Household Size	0.010*	0.011*	0.006	0.006

	(0.006)	(0.006)	(0.006)	(0.006)
Child's Birth-Order	-0.049***	-0.053***	-0.059***	-0.061***
	(0.006)	(0.006)	(0.008)	(0.008)
First Wealth Quintile	-0.280***	-0.280***	-0.277***	-0.276***
	(0.025)	(0.025)	(0.029)	(0.031)
Second Wealth Quintile	-0.138***	-0.140***	-0.136***	-0.138***
	(0.026)	(0.025)	(0.029)	(0.031)
Fourth Wealth Quintile	0.111***	0.112***	0.112***	0.114***
	(0.026)	(0.025)	(0.024)	(0.026)
Fifth Wealth Quintile	0.345***	0.343***	0.340***	0.339***
	(0.031)	(0.031)	(0.033)	(0.035)
Mother's Body Mass Index	0.068***	0.068***	0.068***	0.068***
	(0.003)	(0.003)	(0.003)	(0.003)
Constant	-2.773***	-2.775***	-2.679***	-2.684***
	(0.065)	(0.064)	(0.070)	(0.074)
Observations	20975	20975	20975	20975
Number of Mother_id		14909		14909

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The sample consists of children under age 5 in the DHS-India. Only those children with married mother with at least two children are included.

Table E11. Models for Child's Weight-for-Height Z-Scores -Results from Twins' Data

VARIABLES	(1) OLS	(2) GLS	(3) 2SLS	(4) IV-GLS
Child is female	0.037	0.099	0.554*	0.543**
	(0.176)	(0.172)	(0.322)	(0.239)
Gender Ratio Gap	-0.180	-0.075	-0.754	1.295
	(0.560)	(0.500)	(0.957)	(3.045)
Mother's Bargaining Power	-0.122	-0.172	-0.134	-0.398
	(0.151)	(0.150)	(0.150)	(0.809)
Gender Ratio Gap*Female	-0.501	-0.345	-2.265	-4.450*
	(0.663)	(0.569)	(1.430)	(2.311)
Gender Ratio Gap*Bargaining Power	-0.207	-0.206	-1.297	7.275
	(1.092)	(0.943)	(1.336)	(6.903)
Gender Ratio Gap*Bargaining Power*Female	-0.904	-0.465	-0.312	-4.914
	(1.094)	(0.853)	(1.812)	(3.380)
Age of Child	0.019	0.016	0.005	-0.043
	(0.064)	(0.066)	(0.056)	(0.413)
Household Size	-0.006	0.010	0.022	-0.012
	(0.051)	(0.052)	(0.047)	(0.270)
Child's Birth-Order	-0.015	-0.040	-0.051	-0.055
	(0.042)	(0.040)	(0.048)	(0.060)
First Wealth Quintile	-0.082	-0.071	-0.001	0.253
	(0.309)	(0.318)	(0.259)	(2.134)
Second Wealth Quintile	-0.076	-0.045	0.026	0.311

	(0.246)	(0.245)	(0.271)	(1.959)
Fourth Wealth Quintile	0.508**	0.457*	0.465*	0.171
	(0.245)	(0.245)	(0.240)	(1.907)
Fifth Wealth Quintile	0.820***	0.769***	0.644**	0.113
	(0.265)	(0.259)	(0.272)	(2.453)
Mother's Body Mass Index	0.017	0.018	0.017	0.247
	(0.028)	(0.029)	(0.023)	(0.620)
Constant	-2.529***	-2.578***	-2.577***	-6.786
	(0.722)	(0.723)	(0.586)	(13.233)
Observations	317	317	317	317
Number of Mother_id		152		152

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The sample consists of living twin-pairs (and triplets) under age 5 in the DHS-India. Only those children with married mother with at least two children are included.

Appendix F. Additional Results and Summary Statistics for Essay 3

Table F1. Description and Weighted Summary Statistics by Gender for the Variables Used in the Models

	Boys (N=2972)		Girls (N=3054)	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Schooling Outcome				
High School Completion	0.875	0.331	0.898	0.302
High School Diploma	0.797	0.402	0.838	0.369
College Enrollment	0.492	0.500	0.591	0.492
Enrollment |HS Completion	0.558	0.497	0.657	0.475
Teenage Sex Indicators				
Had Sex Under 15	0.203	0.402	0.160	0.366
Had Sex Under 18	0.504	0.500	0.522	0.500
Sex Partners	5.526	12.436	3.330	5.051
Age at 1st sex	16.926	3.707	17.142	3.212
Total Sex	115.064	264.697	153.572	327.993
Sex Per Partner	2.688	7.013	4.314	10.154
Teenage dating indicators				
Had Date Under 15	0.666	0.472	0.533	0.499
Had Date Under 18	0.937	0.243	0.935	0.247
Dating Partners	13.851	2.462	14.453	2.077
Age at 1st Date	16.673	24.673	10.645	16.058
Total Dates	27.879	84.994	24.006	90.705
Date Per Partner	3.452	13.439	4.443	22.843
Controls for Family Background and demographics				
Black	0.151	0.358	0.161	0.367
Hispanic	0.061	0.239	0.057	0.233
White (excluded)	0.724	0.447	0.721	0.448
Dad college educated	0.533	0.499	0.526	0.499
Mom college educated	0.506	0.500	0.516	0.500
Above mean income	0.474	0.499	0.455	0.498
Rural at age 17	0.279	0.448	0.282	0.450
Control for teenage marriage and cohabitation				
Married under 18	0.114	0.318	0.204	0.403
Cohabited under 18	0.220	0.414	0.359	0.480
Controls for teenage pregnancy and child bearing				
Pregnant under 18 (girls)	-	-	0.216	0.411
Had kids under 18	0.064	0.244	0.201	0.401
Instruments				
Parental church visit1	0.501	0.500	0.508	0.500

Parental church visit2	0.123	0.329	0.119	0.323
Parental church visit3	0.269	0.444	0.258	0.437
Parental church visit4	0.104	0.305	0.112	0.315
Parental church visit5	0.002	0.049	0.003	0.059
Peer church visit	0.257	0.437	0.263	0.440

Source: Various rounds of NLSY97.

Table F2. Description and Weighted Summary Statistics by Race for the Variables Used in the Models

	Black (N=1578)		Hispanic (N=622)		White (N=3480)	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Schooling Outcomes						
High School Completion	0.814	0.389	0.815	0.388	0.906	0.292
High School Diploma	0.740	0.439	0.745	0.436	0.842	0.365
College Enrollment	0.437	0.496	0.458	0.498	0.571	0.495
Enrollment |HS Completion	0.532	0.499	0.562	0.496	0.628	0.483
Teenage Sex Indicators						
Had Sex Under 15	0.323	0.468	0.200	0.400	0.150	0.357
Had Sex Under 18	0.640	0.480	0.562	0.496	0.489	0.500
Sex Partners	5.827	10.587	5.213	11.684	4.085	8.919
Age at 1st sex	15.878	3.539	16.610	3.750	17.266	3.335
Total Sex	82.028	213.841	144.391	323.275	147.127	314.717
Sex Per Partner	1.923	6.890	3.966	11.107	3.832	8.855
Teenage dating indicators						
Had Date Under 15	0.519	0.500	0.578	0.494	0.619	0.486
Had Date Under 18	0.923	0.267	0.934	0.248	0.943	0.232
Dating Partners	14.508	2.455	14.251	2.359	14.048	2.216
Age at 1st Date	12.651	19.461	15.919	26.436	13.701	20.478
Total Dates	17.753	56.116	28.273	106.423	28.079	94.163
Date Per Partner	2.920	10.920	5.051	32.887	4.193	19.244
Controls for Family Background and demographics						
Male	0.481	0.500	0.510	0.500	0.497	0.500
Dad college educated	0.487	0.500	0.351	0.477	0.544	0.498
Mom college educated	0.413	0.492	0.285	0.451	0.544	0.498
Above mean income	0.245	0.430	0.228	0.420	0.527	0.499
Rural at age 17	0.201	0.400	0.068	0.252	0.329	0.470
Control for teenage marriage and cohabitation						
Married under 18	0.075	0.264	0.191	0.393	0.175	0.380
Cohabited under 18	0.243	0.429	0.304	0.460	0.302	0.459
Controls for teenage pregnancy and child bearing						

Pregnant under 18 (girls)	0.328	0.470	0.291	0.454	0.187	0.390
Had kids under 18	0.215	0.411	0.202	0.402	0.111	0.314
Instruments						
Parental church visit1	0.371	0.483	0.454	0.498	0.531	0.499
Parental church visit2	0.169	0.375	0.139	0.346	0.111	0.314
Parental church visit3	0.265	0.441	0.275	0.446	0.265	0.442
Parental church visit4	0.189	0.392	0.091	0.287	0.128	0.334
Parental church visit5	0.006	0.075	0.002	0.046	0.003	0.058
Peer church visit	0.260	0.439	0.219	0.414	0.263	0.440

Source: Various rounds of NLSY97.
Notes: About 5.7% of the youth in sample are from other races. In the regression equations these are included in the excluded category (whites).

Table F3. First Stage Estimates for the Results Reported in Columns 4 of Tables 14 and 17 in the Main Text

	Endogenous Regressor					
VARIABLES	Sex Partners	Age at 1st Sex	Total Sex/100	Dating Partners	Age at 1st Date	Total Dates/10
Church Visit2	-0.794**	0.236***	-0.384***	-0.607	0.136*	0.648
	(0.391)	(0.086)	(0.109)	(0.923)	(0.077)	(0.460)
Church Visit3	-1.107***	0.532***	-0.532***	-0.549	0.329***	0.449
	(0.312)	(0.066)	(0.085)	(0.687)	(0.060)	(0.279)
Church Visit4	-1.111**	0.633***	-0.541***	-1.542*	0.573***	0.228
	(0.445)	(0.085)	(0.107)	(0.796)	(0.080)	(0.343)
Church Visit5	-1.707*	0.894**	-0.869***	-5.525***	0.633	-0.972**
	(0.888)	(0.406)	(0.285)	(1.916)	(0.559)	(0.450)
Peer Church	-0.898***	0.376***	-0.241***	-0.417	0.063	0.744**
	(0.276)	(0.060)	(0.078)	(0.652)	(0.058)	(0.299)
Male	3.194***	-0.532***	-0.225***	7.320***	-0.774***	0.399*
	(0.262)	(0.054)	(0.073)	(0.558)	(0.049)	(0.232)
Black	1.889***	-1.007***	-0.584***	-0.011	0.362***	-0.936***
	(0.347)	(0.071)	(0.079)	(0.680)	(0.061)	(0.268)
Hispanic	0.562	-0.238**	-0.057	1.321	0.161*	0.069
	(0.490)	(0.103)	(0.142)	(1.084)	(0.083)	(0.496)
Dad Col. Educated	-0.796***	0.219***	-0.145*	-0.527	0.141***	-0.223
	(0.284)	(0.059)	(0.081)	(0.595)	(0.052)	(0.238)
Mom Col. Educated	-0.131	0.292***	0.009	0.106	-0.043	-0.048
	(0.288)	(0.059)	(0.084)	(0.611)	(0.054)	(0.243)
Above Mean Income	-1.074***	0.435***	-0.252***	0.910	0.022	0.084
	(0.278)	(0.059)	(0.081)	(0.628)	(0.054)	(0.253)
Rural at Age 17	-0.370	0.145**	0.070	-2.567***	0.327***	-0.124
	(0.290)	(0.063)	(0.090)	(0.615)	(0.057)	(0.268)
Constant	4.382***	15.668***	1.992***	11.160***	14.032***	2.317***
	(0.299)	(0.064)	(0.107)	(0.677)	(0.060)	(0.288)
First Stage F-Stat	[5.70,0.00]	[30.76,0.00]	[12.21,0.00]	[2.33,0.04]	[13.65,0.00]	[3.35,0.01]
χ2(.)-Stat for Sargan	{1.41,0.825}	{1.82,0.77}	{1.41,0.84}	{2.27,0.69}	{9.35,0.053}	{17.16,0}

Test						
Observations	6026	6026	6026	6026	6026	6026

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: The first number in the brackets is the value of F-stat for joint significance of the instruments in the first stage equation and the second number is the associated p-value. The first figure in the set brackets is the value of the Sargan chi-square statistic for testing over identifying restrictions and the second number is the associated p-value.

Table F4. First Stage Estimates for the Results Reported in Columns 5 of Tables 14 and 17 in Main the Text

	Endogenous Regressor					
VARIABLES	Sex Partners	Age at 1st Sex	Total Sex/100	Dating Partners	Age at 1st Date	Total Dates/10
Church Visit2	-0.693*	0.202**	-0.347***	-0.460	0.115	0.630
	(0.390)	(0.085)	(0.107)	(0.923)	(0.076)	(0.464)
Church Visit3	-0.891***	0.452***	-0.431***	-0.232	0.283***	0.386
	(0.315)	(0.065)	(0.083)	(0.688)	(0.060)	(0.279)
Church Visit4	-0.772*	0.531***	-0.462***	-1.052	0.501***	0.218
	(0.449)	(0.084)	(0.106)	(0.806)	(0.080)	(0.350)
Church Visit5	-1.583*	0.828**	-0.748**	-5.335***	0.608	-1.080**
	(0.818)	(0.393)	(0.337)	(1.942)	(0.562)	(0.467)
Peer Church	-0.745***	0.309***	-0.135*	-0.188	0.031	0.662**
	(0.277)	(0.059)	(0.076)	(0.653)	(0.058)	(0.297)
Male	3.427***	-0.654***	0.003	7.676***	-0.821***	0.197
	(0.269)	(0.054)	(0.072)	(0.573)	(0.049)	(0.231)
Black	2.044***	-1.112***	-0.353***	0.235	0.332***	-1.159***
	(0.355)	(0.070)	(0.077)	(0.686)	(0.061)	(0.273)
Hispanic	0.690	-0.280***	-0.015	1.508	0.133	0.052
	(0.490)	(0.101)	(0.139)	(1.082)	(0.083)	(0.492)
Dad Col. Educated	-0.720**	0.187***	-0.098	-0.414	0.125**	-0.258
	(0.283)	(0.058)	(0.078)	(0.593)	(0.052)	(0.240)
Mom Col. Educated	-0.015	0.237***	0.102	0.281	-0.066	-0.125
	(0.287)	(0.058)	(0.082)	(0.613)	(0.054)	(0.242)
Above Mean Income	-0.903***	0.348***	-0.096	1.171*	-0.013	-0.051
	(0.284)	(0.058)	(0.079)	(0.631)	(0.054)	(0.251)
Rural at Age 17	-0.320	0.144**	0.039	-2.500***	0.316***	-0.077
	(0.290)	(0.062)	(0.086)	(0.609)	(0.057)	(0.266)
Married Under 18	-0.285	-0.247***	1.026***	-0.279	0.080	-1.213***
	(0.323)	(0.073)	(0.143)	(0.788)	(0.063)	(0.235)
Cohabited Under 18	2.221***	-0.873***	1.209***	3.285***	-0.465***	-0.841***
	(0.328)	(0.061)	(0.100)	(0.696)	(0.052)	(0.237)
Constant	3.312***	16.173***	1.131***	9.544***	14.251***	3.033***
	(0.367)	(0.073)	(0.105)	(0.754)	(0.069)	(0.312)
First Stage F-Stat	[3.70,0.00]	[21.97,0.00]	[7.94,0.00]	[1.78,0.11]	[10.06,0.00]	[3.02,0.01]
χ2(.)-Stat for Sargan Test	{1.36,0.85}	{1.74,0.78}	{1.84,0.76}	{2.66,0.62}	{8.38,0.08}	{16.07,0.00}
Observations	6026	6026	6026	6026	6026	6026

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The first number in the brackets is the value of F-stat for joint significance of the instruments in the first stage equation and the second number is the associated p-value. The first figure in the set brackets is the value of the Sargan chi-square statistic for testing over identifying restrictions and the second number is the associated p-value.

Table F5. First Stage Estimates for the Results Reported in Columns 6 of Tables 14 and 17 in the Main Text

	Endogenous Regressor					
VARIABLES	Sex Partners	Age at 1st Sex	Total Sex/100	Dating Partners	Age at 1st Date	Total Dates/10
Church Visit2	-0.693*	0.202**	-0.346***	-0.460	0.115	0.630
	(0.391)	(0.085)	(0.106)	(0.923)	(0.077)	(0.464)
Church Visit3	-0.866***	0.431***	-0.412***	-0.256	0.279***	0.355
	(0.316)	(0.065)	(0.082)	(0.688)	(0.060)	(0.279)
Church Visit4	-0.750*	0.513***	-0.445***	-1.074	0.497***	0.190
	(0.450)	(0.084)	(0.105)	(0.808)	(0.080)	(0.349)
Church Visit5	-1.550**	0.801**	-0.722**	-5.367***	0.603	-1.121**
	(0.790)	(0.400)	(0.354)	(1.934)	(0.565)	(0.461)
Peer Church	-0.710**	0.280***	-0.108	-0.220	0.025	0.618**
	(0.279)	(0.059)	(0.075)	(0.656)	(0.058)	(0.297)
Male	3.546***	-0.753***	0.097	7.563***	-0.841***	0.046
	(0.282)	(0.054)	(0.074)	(0.581)	(0.050)	(0.241)
Black	1.930***	-1.018***	-0.443***	0.343	0.352***	-1.016***
	(0.355)	(0.070)	(0.078)	(0.705)	(0.062)	(0.270)
Hispanic	0.645	-0.242**	-0.051	1.551	0.141*	0.109
	(0.490)	(0.101)	(0.138)	(1.082)	(0.083)	(0.492)
Dad Col. Educated	-0.722**	0.189***	-0.100	-0.412	0.125**	-0.255
	(0.283)	(0.057)	(0.078)	(0.593)	(0.052)	(0.240)
Mom Col. Educated	0.021	0.207***	0.131	0.246	-0.073	-0.172
	(0.286)	(0.057)	(0.081)	(0.611)	(0.054)	(0.244)
Above Mean Income	-0.846***	0.300***	-0.050	1.115*	-0.023	-0.125
	(0.288)	(0.058)	(0.079)	(0.637)	(0.054)	(0.255)
Rural at Age 17	-0.299	0.126**	0.056	-2.521***	0.312***	-0.104
	(0.289)	(0.062)	(0.086)	(0.608)	(0.057)	(0.267)
Married Under 18	-0.490	-0.077	0.864***	-0.082	0.115*	-0.953***
	(0.335)	(0.074)	(0.140)	(0.818)	(0.064)	(0.227)
Cohabited Under 18	2.051***	-0.732***	1.075***	3.448***	-0.436***	-0.626**
	(0.318)	(0.062)	(0.103)	(0.709)	(0.053)	(0.250)
Had Kids Under 18	0.964**	-0.799***	0.762***	-0.925	-0.165**	-1.222***
	(0.448)	(0.081)	(0.149)	(0.887)	(0.065)	(0.203)
Constant	3.154***	16.304***	1.007***	9.696***	14.278***	3.233***
	(0.375)	(0.074)	(0.106)	(0.761)	(0.070)	(0.329)
First Stage F-Stat	[3.45,00]	[19.77,0.00]	[7.19,0.00]	[1.82,0.11]	[10.06,0.00]	[2.97,0.01]
χ2(.)-Stat for Sargan Test	{1.58,0.81}	{2.08,0.72}	{1.72,0.79}	{2.27,0.69}	{7.22,0.12}	{15.72,0.00}
Observations	6026	6026	6026	6026	6026	6026

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Notes: The first number in the brackets is the value of F-stat for joint significance of the instruments in the first stage equation and the second number is the associated p-value. The first figure in the set brackets is the value of the Sargan chi-square statistic for testing over identifying restrictions and the second number is the associated p-value.

Table F6. Number of Sex Partners and High School Completion by Age 19- Linear Probability Models

VARIABLES		OLS			2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Sex Partners	-0.003***	-0.003***	-0.002***	-0.047***	-0.049***	-0.046***
	(0.001)	(0.001)	(0.001)	(0.010)	(0.014)	(0.013)
Male	-0.033***	-0.046***	-0.066***	0.108***	0.116**	0.089*
	(0.009)	(0.008)	(0.008)	(0.035)	(0.049)	(0.050)
Black	-0.037***	-0.049***	-0.032***	0.037	0.040	0.045
	(0.011)	(0.011)	(0.011)	(0.026)	(0.033)	(0.030)
Hispanic	-0.025	-0.030*	-0.023	-0.006	-0.002	0.001
	(0.017)	(0.017)	(0.017)	(0.028)	(0.030)	(0.029)
Dad Col. Educated	0.007	0.004	0.004	-0.032*	-0.034*	-0.031
	(0.009)	(0.009)	(0.009)	(0.018)	(0.020)	(0.019)
Mom Col. Educated	0.089***	0.083***	0.076***	0.077***	0.077***	0.073***
	(0.009)	(0.009)	(0.009)	(0.016)	(0.016)	(0.015)
Above Mean Income	0.107***	0.099***	0.089***	0.055***	0.053**	0.049**
	(0.008)	(0.008)	(0.008)	(0.020)	(0.021)	(0.020)
Rural at Age 17	0.035***	0.034***	0.031***	0.016	0.017	0.016
	(0.009)	(0.009)	(0.009)	(0.016)	(0.017)	(0.016)
Had Kids Under 18			-0.158***			-0.112***
			(0.016)			(0.031)
Married Under 18		-0.026**	0.007		-0.041**	-0.016
		(0.013)	(0.013)		(0.021)	(0.021)
Cohabited Under 18		-0.083***	-0.055***		0.028	0.040
		(0.011)	(0.011)		(0.039)	(0.036)
Constant	0.811***	0.855***	0.878***	0.981***	0.987***	0.993***
	(0.011)	(0.011)	(0.011)	(0.043)	(0.043)	(0.039)
Observations	6026	6026	6026	6026	6026	6026

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table F7. Age at First Sex and High School Completion by Age 19- Linear Probability Models

VARIABLES		OLS			2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Age at 1st Sex	0.028***	0.025***	0.022***	0.099***	0.101***	0.095***
	(0.002)	(0.002)	(0.002)	(0.014)	(0.017)	(0.017)
Male	-0.027***	-0.038***	-0.057***	0.011	0.012	-0.001
	(0.008)	(0.008)	(0.008)	(0.012)	(0.014)	(0.016)
Black	-0.017	-0.029**	-0.016	0.048***	0.050**	0.054***
	(0.011)	(0.011)	(0.011)	(0.017)	(0.021)	(0.020)

Hispanic	-0.022	-0.026	-0.020	-0.009	-0.008	-0.006
	(0.017)	(0.017)	(0.017)	(0.019)	(0.019)	(0.019)
Dad Col. Educated	0.003	0.001	0.001	-0.016	-0.016	-0.015
	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
Mom Col. Educated	0.080***	0.076***	0.071***	0.054***	0.054***	0.052***
	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
Above Mean Income	0.097***	0.092***	0.084***	0.063***	0.063***	0.060***
	(0.008)	(0.008)	(0.008)	(0.011)	(0.011)	(0.011)
Rural at Age 17	0.031***	0.031***	0.028***	0.019*	0.019*	0.018*
	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
Had Kids Under 18			-0.142***			-0.080***
			(0.016)			(0.023)
Married Under 18		-0.020	0.009		-0.004	0.012
		(0.013)	(0.013)		(0.014)	(0.014)
Cohabited Under 18		-0.065***	-0.042***		0.007	0.017
		(0.011)	(0.011)		(0.020)	(0.019)
Constant	0.361***	0.443***	0.513***	-0.771***	-0.801***	-0.704**
	(0.041)	(0.042)	(0.042)	(0.218)	(0.275)	(0.290)
Observations	6026	6026	6026	6026	6026	6026

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table F8. Frequency of Teenage Sex and High School Completion by Age 19- Linear Probability Models

VARIABLES		OLS			2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Total Sex/100	-0.010***	-0.008***	-0.006***	-0.113***	-0.120***	-0.115***
	(0.002)	(0.002)	(0.002)	(0.020)	(0.026)	(0.027)
Male	-0.044***	-0.054***	-0.073***	-0.068***	-0.054***	-0.062***
	(0.009)	(0.008)	(0.008)	(0.012)	(0.012)	(0.012)
Black	-0.049***	-0.057***	-0.039***	-0.118***	-0.105***	-0.095***
	(0.011)	(0.011)	(0.011)	(0.019)	(0.018)	(0.020)
Hispanic	-0.028	-0.032*	-0.025	-0.040*	-0.039*	-0.036
	(0.017)	(0.017)	(0.017)	(0.022)	(0.023)	(0.022)
Dad Col. Educated	0.008	0.005	0.005	-0.011	-0.010	-0.009
	(0.009)	(0.009)	(0.009)	(0.013)	(0.013)	(0.013)
Mom Col. Educated	0.089***	0.084***	0.077***	0.085***	0.091***	0.088***
	(0.009)	(0.009)	(0.009)	(0.013)	(0.013)	(0.013)
Above Mean Income	0.108***	0.100***	0.091***	0.077***	0.086***	0.083***
	(0.008)	(0.008)	(0.008)	(0.013)	(0.013)	(0.012)
Rural at Age 17	0.036***	0.035***	0.032***	0.041***	0.038***	0.036***
	(0.009)	(0.009)	(0.009)	(0.013)	(0.013)	(0.013)
Had Kids Under 18			-0.156***			-0.070**
			(0.016)			(0.031)
Married Under 18		-0.018	0.013		0.094***	0.103***
		(0.013)	(0.013)		(0.032)	(0.029)
Cohabited Under 18		-0.079***	-0.053***		0.064*	0.070**
		(0.011)	(0.011)		(0.037)	(0.035)
Constant	0.818***	0.855***	0.876***	1.000***	0.960***	0.965***

	(0.011)	(0.011)	(0.011)	(0.037)	(0.028)	(0.026)
Observations	6026	6026	6026	6026	6026	6026

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table F9. Number of Dating Partners and High School Completion by Age 19- Linear Probability Models

VARIABLES		OLS			2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Dating Partners	-0.000**	-0.000	-0.000*	-0.045**	-0.043**	-0.040**
	(0.000)	(0.000)	(0.000)	(0.018)	(0.021)	(0.020)
Male	-0.038***	-0.052***	-0.071***	0.286**	0.276*	0.230
	(0.009)	(0.009)	(0.008)	(0.134)	(0.166)	(0.153)
Black	-0.042***	-0.054***	-0.036***	-0.052	-0.050	-0.028
	(0.011)	(0.011)	(0.011)	(0.032)	(0.031)	(0.030)
Hispanic	-0.026	-0.031*	-0.024	0.028	0.031	0.035
	(0.017)	(0.017)	(0.017)	(0.055)	(0.057)	(0.054)
Dad Col. Educated	0.010	0.006	0.006	-0.017	-0.013	-0.012
	(0.010)	(0.009)	(0.009)	(0.030)	(0.029)	(0.027)
Mom Col. Educated	0.090***	0.083***	0.076***	0.090***	0.093***	0.084***
	(0.009)	(0.009)	(0.009)	(0.028)	(0.028)	(0.026)
Above Mean Income	0.111***	0.102***	0.092***	0.149***	0.151***	0.135***
	(0.008)	(0.008)	(0.008)	(0.033)	(0.037)	(0.034)
Rural at Age 17	0.035***	0.034***	0.030***	-0.081	-0.073	-0.071
	(0.009)	(0.009)	(0.009)	(0.054)	(0.059)	(0.055)
Had Kids Under 18			-0.161***			-0.196***
			(0.016)			(0.042)
Married Under 18		-0.026**	0.008		-0.040	0.002
		(0.013)	(0.013)		(0.037)	(0.035)
Cohabited Under 18		-0.088***	-0.058***		0.056	0.082
		(0.011)	(0.011)		(0.080)	(0.077)
Constant	0.805***	0.851***	0.875***	1.282***	1.247***	1.250***
	(0.011)	(0.011)	(0.011)	(0.196)	(0.201)	(0.190)
Observations	6026	6026	6026	6026	6026	6026

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table F10. Age at First Date and High School Completion by Age 19- Linear Probability Models

VARIABLES		OLS			2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Age at 1st Date	0.017***	0.015***	0.014***	0.136***	0.132***	0.121***
	(0.002)	(0.002)	(0.002)	(0.025)	(0.029)	(0.028)
Male	-0.029***	-0.042***	-0.062***	0.064***	0.054**	0.028
	(0.009)	(0.009)	(0.008)	(0.022)	(0.026)	(0.026)
Black	-0.049***	-0.060***	-0.042***	-0.101***	-0.106***	-0.086***
	(0.011)	(0.011)	(0.011)	(0.017)	(0.018)	(0.018)
Hispanic	-0.030*	-0.034**	-0.027	-0.054***	-0.053***	-0.045**
	(0.017)	(0.017)	(0.017)	(0.020)	(0.019)	(0.019)

Dad Col. Educated	0.007	0.004	0.004	-0.012	-0.012	-0.011
	(0.009)	(0.009)	(0.009)	(0.012)	(0.012)	(0.012)
Mom Col. Educated	0.090***	0.084***	0.077***	0.091***	0.089***	0.082***
	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
Above Mean Income	0.110***	0.101***	0.091***	0.105***	0.101***	0.092***
	(0.008)	(0.008)	(0.008)	(0.011)	(0.011)	(0.010)
Rural at Age 17	0.030***	0.030***	0.027***	-0.011	-0.007	-0.007
	(0.009)	(0.009)	(0.009)	(0.015)	(0.015)	(0.014)
Had Kids Under 18			-0.158***			-0.138***
			(0.016)			(0.018)
Married Under 18		-0.027**	0.006		-0.041***	-0.010
		(0.013)	(0.013)		(0.015)	(0.015)
Cohabited Under 18		-0.082***	-0.053***		-0.021	-0.001
		(0.011)	(0.011)		(0.020)	(0.019)
Constant	0.563***	0.638***	0.674***	-1.132***	-1.044**	-0.871**
	(0.034)	(0.035)	(0.035)	(0.354)	(0.418)	(0.410)
Observations	6026	6026	6026	6026	6026	6026

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table F11. Frequency of Teenage Dating and High School Completion by Age 19- Linear Probability Models

VARIABLES		OLS			2SLS	
	(1)	(2)	(3)	(4)	(5)	(6)
Total Dates	0.002***	0.001***	0.001***	0.058***	0.053***	0.048**
	(0.000)	(0.000)	(0.000)	(0.019)	(0.019)	(0.019)
Male	-0.043***	-0.055***	-0.074***	-0.065***	-0.065***	-0.076***
	(0.009)	(0.008)	(0.008)	(0.018)	(0.016)	(0.014)
Black	-0.041***	-0.053***	-0.035***	0.008	0.004	0.009
	(0.011)	(0.011)	(0.011)	(0.022)	(0.024)	(0.022)
Hispanic	-0.027	-0.032*	-0.025	-0.034	-0.037	-0.032
	(0.017)	(0.017)	(0.017)	(0.033)	(0.030)	(0.028)
Dad Col. Educated	0.010	0.006	0.006	0.019	0.017	0.016
	(0.010)	(0.009)	(0.009)	(0.018)	(0.017)	(0.016)
Mom Col. Educated	0.090***	0.083***	0.076***	0.089***	0.087***	0.082***
	(0.009)	(0.009)	(0.009)	(0.017)	(0.016)	(0.015)
Above Mean Income	0.111***	0.101***	0.091***	0.102***	0.102***	0.095***
	(0.008)	(0.008)	(0.008)	(0.017)	(0.015)	(0.015)
Rural at Age 17	0.036***	0.035***	0.031***	0.040**	0.037**	0.035**
	(0.009)	(0.009)	(0.009)	(0.018)	(0.017)	(0.016)
Had Kids Under 18			-0.159***			-0.099***
			(0.016)			(0.030)
Married Under 18		-0.024*	0.009		0.039	0.054**
		(0.013)	(0.013)		(0.026)	(0.022)
Cohabited Under 18		-0.088***	-0.059***		-0.040	-0.026
		(0.011)	(0.011)		(0.025)	(0.021)
Constant	0.796***	0.843***	0.868***	0.647***	0.670***	0.702***
	(0.011)	(0.011)	(0.011)	(0.051)	(0.065)	(0.067)
Observations	6026	6026	6026	6026	6026	6026

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table F12. First Stage Estimates for the Results Reported in Column 4 of Table 20 in the Main Text

	Endogenous Regressor					
VARIABLES	Sex Partners	Age at 1st Sex	Total Sex/100	Dating Partners	Age at 1st Date	Total Dates/10
Church Visit2	-0.240	0.205**	-0.318***	0.043	0.136*	0.662
	(0.420)	(0.090)	(0.113)	(0.967)	(0.083)	(0.524)
Church Visit3	-1.022***	0.513***	-0.479***	-0.488	0.327***	0.409
	(0.304)	(0.067)	(0.085)	(0.697)	(0.064)	(0.304)
Church Visit4	-1.176***	0.598***	-0.502***	-1.917**	0.583***	0.230
	(0.413)	(0.089)	(0.107)	(0.794)	(0.085)	(0.390)
Church Visit5	-1.399	0.769*	-0.824***	-5.408***	0.597	-1.131**
	(0.897)	(0.414)	(0.285)	(1.898)	(0.562)	(0.463)
Peer Church	-1.016***	0.391***	-0.154*	-0.343	0.054	0.851**
	(0.256)	(0.060)	(0.079)	(0.656)	(0.061)	(0.331)
Male	2.521***	-0.440***	-0.240***	6.563***	-0.722***	0.366
	(0.265)	(0.057)	(0.073)	(0.578)	(0.053)	(0.260)
Black	2.039***	-0.999***	-0.434***	0.350	0.334***	-0.948***
	(0.365)	(0.076)	(0.082)	(0.724)	(0.066)	(0.306)
Hispanic	0.729	-0.308***	-0.069	0.645	0.189**	0.266
	(0.512)	(0.118)	(0.143)	(1.048)	(0.093)	(0.597)
Dad Col. Educated	-0.989***	0.204***	-0.094	-0.446	0.158***	-0.277
	(0.275)	(0.062)	(0.083)	(0.616)	(0.057)	(0.272)
Mom Col. Educated	0.223	0.233***	-0.015	0.287	-0.109*	-0.103
	(0.283)	(0.062)	(0.085)	(0.627)	(0.058)	(0.277)
Above Mean Income	-0.738***	0.309***	-0.121	0.868	-0.058	0.016
	(0.273)	(0.061)	(0.081)	(0.643)	(0.057)	(0.277)
Rural at Age 17	-0.271	0.086	-0.006	-2.506***	0.338***	-0.265
	(0.281)	(0.066)	(0.087)	(0.641)	(0.060)	(0.292)
Constant	4.044***	15.876***	1.770***	11.059***	14.132***	2.551***
	(0.313)	(0.071)	(0.110)	(0.728)	(0.066)	(0.342)
First Stage F-Stat	[7.34,0.00]	[27.44,0.00]	[9.15,0.00]	[2.73,0.02]	[12.12,0.00]	[3.54,0.00]
$\chi2(.)$-Stat for Sargan Test	{6.71,0.15}	{10.78,0.00}	{9.14,0.06}	{22.68,0.00}	{28.61,0.00}	{3.76,0.44}
Observations	5208	5208	5208	5208	5208	5208
R-squared	0.039	0.096	0.017	0.029	0.061	0.005

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The first number in the brackets is the value of F-stat for joint significance of the instruments in the first stage equation and the second number is the associated p-value. The first figure in the set brackets is the value of the Sargan chi-square statistic for testing over identifying restrictions and the second number is the associated p-value.

Table F13. First Stage Estimates for the Results Reported in Column 5 of Table 20 in the Main Text

	Endogenous Regressor					
VARIABLES	Sex Partners	Age at 1st Sex	Total Sex/100	Dating Partners	Age at 1st Date	Total Dates/10
Church Visit2	-0.155	0.171*	-0.278**	0.188	0.118	0.645
	(0.419)	(0.089)	(0.110)	(0.967)	(0.082)	(0.528)
Church Visit3	-0.812***	0.430***	-0.384***	-0.133	0.281***	0.370
	(0.312)	(0.067)	(0.083)	(0.698)	(0.064)	(0.304)
Church Visit4	-0.874**	0.495***	-0.430***	-1.426*	0.511***	0.280
	(0.420)	(0.088)	(0.107)	(0.803)	(0.086)	(0.397)
Church Visit5	-1.293	0.716*	-0.731**	-5.213***	0.579	-1.223**
	(0.833)	(0.402)	(0.332)	(1.918)	(0.565)	(0.481)
Peer Church	-0.863***	0.323***	-0.056	-0.074	0.025	0.777**
	(0.258)	(0.059)	(0.077)	(0.656)	(0.061)	(0.329)
Male	2.750***	-0.551***	-0.052	6.977***	-0.762***	0.189
	(0.273)	(0.057)	(0.071)	(0.593)	(0.053)	(0.258)
Black	2.178***	-1.081***	-0.263***	0.619	0.317***	-1.148***
	(0.367)	(0.075)	(0.081)	(0.725)	(0.067)	(0.310)
Hispanic	0.840*	-0.351***	-0.022	0.833	0.165*	0.251
	(0.510)	(0.116)	(0.141)	(1.044)	(0.092)	(0.592)
Dad Col. Educated	-0.923***	0.177***	-0.059	-0.332	0.145**	-0.298
	(0.274)	(0.061)	(0.080)	(0.613)	(0.057)	(0.274)
Mom Col. Educated	0.317	0.188***	0.059	0.455	-0.126**	-0.170
	(0.282)	(0.060)	(0.082)	(0.629)	(0.057)	(0.275)
Above Mean Income	-0.594**	0.239***	-0.003	1.128*	-0.083	-0.097
	(0.278)	(0.060)	(0.079)	(0.643)	(0.057)	(0.274)
Rural at Age 17	-0.235	0.082	-0.027	-2.457***	0.326***	-0.210
	(0.283)	(0.065)	(0.083)	(0.636)	(0.060)	(0.291)
Married Under 18	-0.026	-0.209***	0.856***	0.228	0.105	-1.400***
	(0.352)	(0.079)	(0.141)	(0.844)	(0.068)	(0.236)
Cohabited Under 18	2.122***	-0.881***	1.142***	3.647***	-0.436***	-0.679**
	(0.331)	(0.067)	(0.105)	(0.735)	(0.057)	(0.283)
Constant	3.047***	16.337***	1.046***	9.286***	14.316***	3.172***
	(0.360)	(0.078)	(0.106)	(0.785)	(0.074)	(0.357)
First Stage F-Stat	[4.62,0.00]	[19.16,0.00]	[5.91,0.00]	[2.12,0.06]	[8.92,0.00]	[3.29,0.01]
$\chi 2(.)$-Stat for Sargan Test	{6.39,0.17}	{9.77,0.04}	{10.17,0.04}	{44.27,0.00}	{24.50,0.00}	{3.43,0.49}
Observations	5208	5208	5208	5208	5208	5208
R-squared	0.048	0.129	0.066	0.035	0.070	0.009

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: The first number in the brackets is the value of F-stat for joint significance of the instruments in the first stage equation and the second number is the associated p-value. The first figure in the set brackets is the value of the Sargan chi-square statistic for testing over identifying restrictions and the second number is the associated p-value.

Table F14. First Stage Estimates for the Results Reported in Column 6 of Table 20 in the Main Text

	Endogenous Regressor					
VARIABLES	Sex Partners	Age at 1st Sex	Total Sex/100	Dating Partners	Age at 1st Date	Total Dates/10
Church Visit2	-0.147	0.166*	-0.274**	0.189	0.117	0.637
	(0.420)	(0.088)	(0.110)	(0.967)	(0.082)	(0.527)
Church Visit3	-0.783**	0.413***	-0.370***	-0.132	0.278***	0.341
	(0.313)	(0.066)	(0.083)	(0.698)	(0.064)	(0.304)
Church Visit4	-0.845**	0.477***	-0.416***	-1.425*	0.507***	0.251
	(0.422)	(0.087)	(0.107)	(0.806)	(0.086)	(0.396)
Church Visit5	-1.267	0.701*	-0.719**	-5.212***	0.576	-1.248***
	(0.796)	(0.409)	(0.346)	(1.919)	(0.567)	(0.477)
Peer Church	-0.821***	0.298***	-0.036	-0.073	0.019	0.735**
	(0.258)	(0.059)	(0.077)	(0.658)	(0.061)	(0.329)
Male	2.878***	-0.629***	0.009	6.982***	-0.777***	0.062
	(0.289)	(0.057)	(0.073)	(0.602)	(0.054)	(0.268)
Black	2.027***	-0.989***	-0.335***	0.614	0.335***	-0.998***
	(0.365)	(0.075)	(0.080)	(0.744)	(0.067)	(0.308)
Hispanic	0.776	-0.312***	-0.052	0.831	0.173*	0.315
	(0.514)	(0.116)	(0.140)	(1.048)	(0.092)	(0.593)
Dad Col. Educated	-0.931***	0.182***	-0.063	-0.332	0.146***	-0.290
	(0.274)	(0.061)	(0.080)	(0.613)	(0.056)	(0.274)
Mom Col. Educated	0.351	0.167***	0.076	0.456	-0.130**	-0.204
	(0.282)	(0.060)	(0.082)	(0.626)	(0.057)	(0.277)
Above Mean Income	-0.536*	0.204***	0.025	1.130*	-0.090	-0.155
	(0.282)	(0.060)	(0.080)	(0.649)	(0.057)	(0.277)
Rural at Age 17	-0.212	0.067	-0.016	-2.456***	0.323***	-0.234
	(0.281)	(0.065)	(0.083)	(0.635)	(0.060)	(0.292)
Married Under 18	-0.274	-0.058	0.737***	0.219	0.134*	-1.154***
	(0.363)	(0.081)	(0.138)	(0.866)	(0.069)	(0.225)
Cohabited Under 18	1.945***	-0.773***	1.058***	3.641***	-0.415***	-0.504*
	(0.317)	(0.068)	(0.109)	(0.746)	(0.058)	(0.295)
Had Kids Under 18	1.277**	-0.774***	0.610***	0.045	-0.154**	-1.260***
	(0.547)	(0.094)	(0.162)	(1.003)	(0.076)	(0.221)
Constant	2.881***	16.437***	0.967***	9.280***	14.336***	3.336***
	(0.365)	(0.078)	(0.108)	(0.789)	(0.075)	(0.371)
First Stage F-Stat	[4.15,0.00]	[17.35,0.00]	[5.51,0.00]	[2.11,0.06]	[8.74,0.00]	[3.22,0.01]
$\chi 2(.)$-Stat for Sargan Test	{6.65,0.16}	{10.06,0.04}	{10.11,0.04}	{40.69,0.00}	{23.28,0.00}	{3.49,0.48}
Observations	5208	5208	5208	5208	5208	5208

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: The first number in the brackets is the value of F-stat for joint significance of the instruments in the first stage equation and the second number is the associated p-value. The first figure in the set brackets is the value of the Sargan chi-square statistic for testing over identifying restrictions and the second number is the associated p-value.

Table F15. Linear Probability Models for College Enrollment at or Before age 20 Conditional on High School Completion (Girls)

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.012***	-0.009***	-0.008***	-0.097***	-0.102***	-0.088**
	(0.003)	(0.002)	(0.002)	(0.024)	(0.033)	(0.035)
Age at 1st sex	0.043***	0.031***	0.021***	0.126***	0.119***	0.092**
	(0.006)	(0.005)	(0.005)	(0.028)	(0.034)	(0.039)
Total sex /100	-0.019***	-0.010***	-0.007**	-0.169***	-0.167***	-0.156***
	(0.003)	(0.003)	(0.003)	(0.042)	(0.053)	(0.060)
Dating Partners	-0.002***	-0.002***	-0.002***	-0.019	0.018	0.017
	(0.001)	(0.000)	(0.001)	(0.016)	(0.021)	(0.020)
Age at 1st date	0.013***	0.008	0.004	0.166***	0.151**	0.105
	(0.005)	(0.005)	(0.005)	(0.053)	(0.066)	(0.066)
Total Dates/10	0.000	-0.000	-0.001	0.086*	0.064	0.053
	(0.001)	(0.001)	(0.001)	(0.050)	(0.042)	(0.039)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	2696	2696	2696	2696	2696	2696

Standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the sex and dating experience indicators were separately entered the regression equations.

Table F16. Linear Probability Models for College Enrollment at or Before age 20 Conditional on High School Completion (Boys)

	OLS			2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.004***	-0.003***	-0.003***	-0.048***	-0.046***	-0.047***
	(0.001)	(0.001)	(0.001)	(0.015)	(0.016)	(0.017)
Age at 1st sex	0.038***	0.034***	0.034***	0.177***	0.176***	0.176***
	(0.004)	(0.004)	(0.004)	(0.036)	(0.041)	(0.041)
Total sex /100	-0.018***	-0.011***	-0.011***	-0.249***	-0.244***	-0.241***
	(0.004)	(0.004)	(0.004)	(0.073)	(0.087)	(0.086)
Dating Partners	-0.001***	-0.001***	-0.001***	-0.030*	-0.019	-0.019
	(0.000)	(0.000)	(0.000)	(0.017)	(0.014)	(0.013)
Age at 1st date	0.017***	0.015***	0.015***	0.181***	0.170***	0.170***
	(0.005)	(0.004)	(0.004)	(0.054)	(0.055)	(0.055)
Total Dates/10	0.001	0.000	0.000	0.073***	0.068**	0.068**
	(0.001)	(0.001)	(0.001)	(0.027)	(0.027)	(0.027)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes

Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	2512	2512	2512	2512	2512	2512

Standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the sex and dating experience indicators were separately entered the regression equations.

Table F17. Marginal Effects from Probit Models for High School Completion by Age 19

	Probit			IV-Probit[aaa]		
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.002***	-0.002***	-0.001***	-0.035***	-0.037***	-0.036***
	(0.000)	(0.000)	(0.000)	(0.003)	(0.003)	(0.003)
Age at 1st sex	0.021***	0.018***	0.016***	0.110***	0.115***	0.114***
	(0.002)	(0.002)	(0.002)	(0.014)	(0.017)	(0.018)
Total sex /100	-0.008***	-0.006***	-0.004***	-0.107***	-0.114***	-0.113***
	(0.001)	(0.001)	(0.002)	(0.011)	(0.012)	(0.013)
Dating Partners	-0.000**	-0.000*	-0.000**	-0.018***	-0.019***	-0.019***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Age at 1st date	0.015***	0.013***	0.012***	0.154***	0.157***	0.151***
	(0.002)	(0.002)	(0.002)	(0.018)	(0.022)	(0.024)
Total Dates/10	0.003**	0.002**	0.002**	0.043***	0.043***	0.043***
	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.004)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	6026	6026	6026	6026	6026	6026

Standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the teenage sex and dating experience indicators were separately entered the regression equations. [aaa]The IV-Probit models assumes continuity in the endogenous regressor which does not strictly hold for the indicators of teenage dating/sex whose marginal effects are reported in this tables.

Table F18. Marginal Effects from Probit Models for High School Completion by Age 19 (Girls)

	Probit			IV-Probit[aaa]		
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.003**	-0.002**	-0.001	-0.065***	-0.070***	-0.066***
	(0.001)	(0.001)	(0.001)	(0.009)	(0.010)	(0.013)
Age at 1st sex	0.022***	0.018***	0.011***	0.095***	0.099***	0.082**
	(0.003)	(0.003)	(0.002)	(0.021)	(0.028)	(0.034)
Total sex /100	-0.006***	-0.004**	-0.001	-0.099***	-0.106***	-0.100***
	(0.001)	(0.001)	(0.001)	(0.015)	(0.018)	(0.025)

Dating Partners	0.000	0.000	0.000	-0.024***	0.025***	0.025***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Age at 1st date	0.013***	0.010***	0.006**	0.155***	0.167***	0.122
	(0.003)	(0.003)	(0.003)	(0.035)	(0.051)	(0.075)
Total Dates/10	0.009**	0.007**	0.004*	0.046***	0.045***	0.045***
	(0.004)	(0.003)	(0.003)	(0.005)	(0.005)	(0.006)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	2696	2696	2696	2696	2696	2696

Standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the teenage sex and dating experience indicators were separately entered the regression equations. [aaa]The IV-Probit models assumes continuity in the endogenous regressor which does not strictly hold for the indicators of teenage dating/sex whose marginal effects are reported in this tables.

Table F19. Marginal Effects from Probit Models for High School Completion by Age 19 (Boys)

	Probit			IV-Probit[aaa]		
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.002***	-0.002***	-0.002***	-0.027***	-0.028***	-0.028***
	(0.000)	(0.000)	(0.000)	(0.003)	(0.003)	(0.003)
Age at 1st sex	0.0195***	0.0181***	0.0175***	0.115***	0.119***	0.119***
	(0.00241)	(0.00239)	(0.00239)	(0.0203)	(0.0223)	(0.0221)
Total sex /100	-0.010***	-0.008***	-0.007***	-0.120***	-0.128***	-0.128***
	(0.002)	(0.002)	(0.002)	(0.016)	(0.017)	(0.017)
Dating Partners	-0.001***	-0.000**	-0.001**	-0.015***	-0.015***	-0.015***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Age at 1st date	0.016***	0.015***	0.014***	0.156***	0.155***	0.156***
	(0.003)	(0.003)	(0.003)	(0.022)	(0.024)	(0.024)
Total Dates/10	0.002	0.001	0.001	0.040***	0.040***	0.040***
	(0.001)	(0.001)	(0.001)	(0.005)	(0.005)	(0.005)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	2696	2696	2696	2696	2696	2696

Standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the teenage sex and dating experience indicators were separately entered the regression equations. [aaa]The IV-Probit models assumes continuity in the endogenous regressor which does not strictly hold for the indicators of teenage dating/sex whose marginal effects are reported in this tables.

Table F20. Marginal Effects from Probit Models for College Enrollment at or Before age 20 Conditional on High School Completion

	Probit			IV-Probit[aaa]		
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.005***	-0.004***	-0.004***	-0.041***	-0.042***	-0.041***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
Age at 1st sex	0.047***	0.039***	0.035***	0.150***	0.148***	0.145***
	(0.004)	(0.004)	(0.004)	(0.013)	(0.016)	(0.018)
Total sex /100	-0.018***	-0.009***	-0.007**	-0.136***	-0.139***	-0.137***
	(0.003)	(0.003)	(0.003)	(0.009)	(0.011)	(0.012)
Dating Partners	-0.001***	-0.001	-0.001	-0.019***	-0.020***	-0.020***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Age at 1st date	0.023***	0.019***	0.018***	0.179***	0.176***	0.169***
	(0.004)	(0.004)	(0.004)	(0.017)	(0.024)	(0.029)
Total Dates/10	0.001	0.000	-0.000	0.041***	0.040***	0.040***
	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.004)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	5208	5208	5208	5208	5208	5208

Standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the teenage sex and dating experience indicators were separately entered the regression equations. [aaa]The IV-Probit models assumes continuity in the endogenous regressor which does not strictly hold for the indicators of teenage dating/sex whose marginal effects are reported in this tables.

Table F21. Marginal Effects from Probit Models for College Enrollment at or Before age 20 Conditional on High School Completion (Girls)

	Probit			IV-Probit[aaa]		
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.009**	-0.007**	-0.004	-0.070***	-0.071***	-0.066***
	(0.004)	(0.003)	(0.003)	(0.009)	(0.012)	(0.016)
Age at 1st sex	0.0503***	0.040***	0.027***	0.124***	0.114***	0.084*
	(0.006)	(0.006)	(0.006)	(0.026)	(0.036)	(0.047)
Total sex /100	-0.017***	-0.008**	-0.004	-0.116***	-0.115***	-0.107***
	(0.004)	(0.004)	(0.004)	(0.015)	(0.022)	(0.033)
Dating Partners	-0.001	-0.001	-0.001	-0.025***	0.025***	0.025***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
Age at 1st date	0.021***	0.016***	0.010*	0.172***	0.167**	0.096
	(0.005)	(0.006)	(0.006)	(0.038)	(0.071)	(0.136)
Total Dates/10	0.001	0.000	0.000	0.041***	0.040***	0.039***
	(0.001)	(0.001)	(0.001)	(0.006)	(0.006)	(0.007)
Control Vars.						
Background &						

Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	2696	2696	2696	2696	2696	2696

Standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the teenage sex and dating experience indicators were separately entered the regression equations. [aaa]The IV-Probit models assumes continuity in the endogenous regressor which does not strictly hold for the indicators of teenage dating/sex whose marginal effects are reported in this tables.

Table F22. Marginal Effects from Probit Models for College Enrollment at or Before age 20 Conditional on High School Completion (Boys)

		Probit			IV-Probit[aaa]	
	(1)	(2)	(3)	(4)	(5)	(6)
Sex partners	-0.005***	-0.004***	-0.004***	-0.032***	-0.033***	-0.033***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.0023)
Age at 1st sex	0.045***	0.040***	0.039***	0.161***	0.161***	0.161***
	(0.005)	(0.005)	(0.005)	(0.013)	(0.015)	(0.015)
Total sex /100	-0.018***	-0.010**	-0.009*	-0.159***	-0.163***	-0.164***
	(0.004)	(0.005)	(0.005)	(0.011)	(0.012)	(0.012)
Dating Partners	-0.001**	-0.001	-0.001	-0.016***	-0.016***	-0.016***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Age at 1st date	0.024***	0.021***	0.021***	0.184***	0.180***	0.180***
	(0.005)	(0.005)	(0.005)	(0.015)	(0.019)	(0.019)
Total Dates/10	0.000	-0.000	-0.000	0.041***	0.041***	0.041***
	(0.001)	(0.001)	(0.001)	(0.004)	(0.004)	(0.004)
Control Vars.						
Background & Demographics	Yes	Yes	Yes	Yes	Yes	Yes
Cohabitation & Marriage	No	Yes	Yes	No	Yes	Yes
Teenage Children	No	No	Yes	No	No	Yes
Observations	2512	2512	2512	2512	2512	2512

Standard errors in parentheses

***Significant at 1%; ** significant at 5%; * significant at 10%

Notes: Each of the teenage sex and dating experience indicators were separately entered the regression equations. [aaa]The IV-Probit models assumes continuity in the endogenous regressor which does not strictly hold for the indicators of teenage dating/sex whose marginal effects are reported in this tables.

REFERENCES

Abadian, Sousan.1996. Women's autonomy and its impact on fertility. *World Development* 24(12):1793-1809.

Adato, M., B. de la Briere, D. Mindek, and A.R. Quisumbing. 2003. The impact of PROGRESA on women's status and intrahousehold relations. In A.R. Quisumbing, ed., *Household decisions, gender, and development: A synthesis of recent research*. Washington, D.C.: International Food Policy Research Institute.

Ahmed, Meherun. 2006. Intra-household bargaining and investment in child health. Unpublished preliminary report. Northfield, Minnesota: Department of Economics, Carlton College,

Ahn, Namkee. 1994. Teenage childbearing and high school completion: Accounting for individual heterogeneity. *Family Planning Perspectives* 26(1):17-21.

Alderman, H., J. Behrman, V. Lavy, and R. Menon. 2001. 'Child health and school enrollment: A longitudinal analysis'. *Journal of Human Resources* 36: 185-205.

Alderman, H., J. Hoddinott, and B. Kinsey. 2006. Long term consequences of early childhood malnutrition. *Oxford Economic Papers* 58(3):450-474.

Angrist, J., and W. Evans. 1998. Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review* 88(3):450-477.

Angrist, J.D., and Allan Krueger. 2001. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives* 15(4): 69–85.

Ashenfelter, O., and Allan Krueger. 1994. Estimates of the economic returns to schooling from a new sample of twins. *American Economic Review* 84(5):1157-1173.

Bacolod, Marigee P., and Priya Ranjan. 2008. Why children work, attend school, or stay idle: The roles of ability and household wealth. *Economic Development and Cultural Change* 56:791-828.

Baez, Javier E. 2008. Does More Mean Better? Sibling sex composition and the link between family size and children's quality. IZA Discussion Papers No. 3472.

Barro, Robert J., and Gary S. Becker. 1989. Fertility choice in a model of economic growth. *Econometrica* 57(2):481-501.

Baser, O., C.J. Bradley, J.C. Gardiner, and C. Given. 2004. Testing and correcting for non-random selection bias due to censoring: An application to medical costs. *Health Services and Outcomes Research Methodology* 4: 93–107.

Basmann, R.L. 1960. On finite sample distributions of generalized classical linear identifiability test statistics. *Econometrica* 45:939-952.

Basu, Kaushik. 2006. Gender and Say: A model of household behavior with endogenously-determined balance of power. *The Economic Journal* 116(511):558–80.

Becker, Gary S. 1960. An economic analysis of fertility. In *demographic and economic change in developed countries*. Universities-National Bureau Conference Series, No.11. Princeton, NJ. : Princeton University Press.

____________. 1980. *A Treatise on the family*. Cambridge and London: Harvard University Press.

____________. 1991. *A Treatise on the family*. Cambridge and London: Harvard University Press.

Becker, Gary S., and Gregg H. Lewis. 1973. On the interaction between quality and quantity of children. *Journal of Political Economy* 84: S279-S288.

Becker, Gary S., and Nigel Tomes. 1976. Child endowments and the quality and quantity of children. *Journal of Political Economy*, 82: S143-S162.

Behrman, Jere R., and Victor Lavy. 1994. Children's health and achievement in school. LSMS Living Standards Measurement Study Working Paper 104.

Behrman, Jere R., and Mark R. Rosenzweig. 2002. Does increasing women's schooling raise the schooling of the next generation?' *American Economic Review* 92(1): 323-334.

Behrman, Jere R., Mark R. Rosenzweig, and Paul Taubman. 1994. Endowments and the allocation of schooling in the family and in the marriage market: The twins experiment. *Journal of Political Economy* 102(6):1131-1174.

Billy, J. O. G., N. S. Landale, W. R. Grady, and D. D. Zimmerle. 1988. Effects of sexual activity on adolescent social and psychological development. *Social Science Quarterly* 51(3):190-212.

Binder, M. 1999. Schooling indicators during Mexico's 'lost decade.' *Economics of Education Review* 18(2):183–99.

Black, Sandra E., Paul J. Devereux and Kjell G. Salvanes. 2007. Small family, smart family? The family size and IQ scores of young men. NBER Working Paper 13336.

Bourguignon, F., F. Ferreira, and P. Leite. 2003. Conditional cash transfers, schooling, and child labor: Micro-simulating Brazil's Bolsa Escola Program. *World Bank Economic Review* 17(2):229–54.

Brewster, Karin L., Elizabeth C. Cooksey, David K. Guilkey, and Ronald R. Rindfuss. 1998. The changing impact of religion on the sexual and contraceptive behavior of adolescent women in the United States. *Journal of Marriage and the Fam*ily 60(2): 493-504.

Bronars, S., and J. Grogger. 1994. The economic consequences of unwed motherhood: Using twin births as a natural experiment. *American Economic Review* 84(5):1141-56.

Browning, M., and P.A. Chiappori. 1998. Efficient intra-household allocations: A general characterizations and empirical tests. *Econometrica* 66(6):1241-1278.

Butcher, Kristin F., and Case Anne. 1994. The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics* 109(3):531-63.

Cameron, S., and J. Heckman. 1999. Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. *Journal of Political Economy* 06(2):262-333.

___________. 2001. The Dynamics of educational attainment for black, Hispanic, and white Males. *Journal of Political Economy* 109:455–99.

Cameron, Stephen V., and Christopher Taber. 2000. Borrowing constraints and the returns to schooling. NBER Working Paper No. w7761.

___________. 2004. Estimation of educational borrowing constraints using returns to schooling. *Journal of Political Economy* 112(1):132-182.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and applications*, Cambridge.

Carlin, John B., Lyle C. Gurrin, Jonathan AC Sterne, Ruth Morley, and Terry Dwyer. 2005. Regression models for twin studies: A critical review. *International Journal of Epidemiology* 34:1089-1099.

Carneiro, Pedro, and James J. Heckman. 2002. The evidence on credit constraints in post-secondary schooling. *The Economic Journal* 112(482):705-734.

Carter, R. Michael., and A. John Maluccio. 2003. Social capital and coping with economic shocks: An analysis of stunting of South African children. *World Development* 31(7):1147-1163.

Chatterjia, Pinka. 2006. Illicit drug use and educational attainment. *Health Economics* 15:489-511.

Chiappori, P.A. 1992. Collective labor supply and welfare. *Journal of Political Economy* 100:437-467.

_________.1997. Introducing household production in collective models of labor supply. *Journal of Political Economy* 105(1):191-209.

Cigno, Allessandro, and Furio Rosati. 2005. *The economics of child labor*. New York: Oxford University Press.

Conley, Dalton. 2000. Sibling sex composition: Effects on educational attainment. *Social Science Research* 29:441-457.

Conley, Dalton, Kate W. Strully, and Neil G. Bennet. 2006. Twin differences in birth weight: The effect of genotype and prenatal environment on neonatal and post-neonatal mortality. *Economics and Human Biology* 4(2):151-183.

Dammert, Ana. 2008. Child labor and schooling responses to changes in coca production in rural Peru. *Journal of Development Economics* 86:164-180.

Dasgupta, Partha. 1997. Nutritional status, the capacity for work and poverty traps. *Journal of Econometrics* 77(1): 5-38.

De Luca, Giuseppe. 2008. SNP and SML estimation of univariate and bivariate binary-choice models. *The State Journal* 8(2):190–220.

Dercon, Stefan. 2004. Growth and shocks: Evidence from rural Ethiopia. *Journal of Development Economics* 74:309-329.

_________. 2000. Changes in poverty and social indicators in Ethiopia in the 1990s. Oxford University: Centre for the Study of African Economies (CSAE).

Dercon, S., and P. Krishnan. 2000. Vulnerability, seasonality and poverty in Ethiopia. *Journal of Development Studies* 36(6):25-53.

Desai, Sonalde, and Soumya Alava. 1998. Maternal education and child health: Is there a strong causal relationship? *Demography* 35(1):71-81.

DeSimone, Jeff. 2002. Illegal drug use and employment. *Journal of Labor Economics,* 20:952–977.

DeSimone, Jeff, and Amy M. Wolaver. 2005. Drinking and academic performance in high school. NBER Working Paper No. 11035.

Dillon, Andrew. 2008. Child labor and schooling responses to production and health shocks in northern Mali. IFPRI Discussion Paper 00755.

Duflo, E. 2003. Grandmothers and granddaughters: Old-age pensions and intra-household allocation in South Africa. *World Bank Economic Review* 17(1):1–25.

Edmonds, Eric. 2006. Child labor and schooling responses to anticipated income in South Africa. *Journal of Development Economics* 81(2):386–414.

Emerson, P., and A.P. Souza. 2007. Child labor, school attendance, and intrahousehold gender bias in Brazil. *World Bank Economic Review* 21(2):301–16.

Evans, William N., Wallace E. Oates, and Robert M. Schwab. 1992. Measuring peer group effects: A study of teenage behavior. *The Journal of Political Economy* 100(5):966-991.

Farré, L., R. Klein, and F. Vella. 2009. Does increasing parents' schooling raise the schooling of the next generation? Evidence based on conditional second moments. IZA Discussion paper 3967.

Fletcher, Jason M. 2009. Beauty vs. brains: Early labor market outcomes of high school graduates. *Economics Letters* 105:321-325.

Garg, Ashish, and Jonathan Morduch. 1998. Siblings rivalry and the gender gap: Evidence from child health outcomes from Ghana. *Journal of Population Economics* 11:461-493.

Geronimus, A. T., and S. Korenman. 1992. The socioeconomic consequences of teen childbearing reconsidered. *Quarterly Journal of Economics* 107:1187-1214.

Gibson, Mhairi A. 2008. Does Investment in the sexes differ when fathers are absent? Sex-biased infant survival and child growth in rural Ethiopia. *Human Nature* 19(3):263-276.

Gibson, T.B., T.L. Mark, K. Axelsen, O. Baser, D.A. Rublee, and K.A. McGuigan. 2006. Impact of statin copayments on adherence and medical care utilization and expenditures. *American Journal of Managed Care* 12: SP11–SP19.

Gitter , Seth R., and Bradford L. Barham. 2008. Women's power, conditional cash transfers, and schooling in Nicaragua. *The World Bank Economic Review* 22(2):271-290.

Glewwe, Paul .2002. Schools and skills in developing countries: Education policies and socioeconomic outcomes Source. *Journal of Economic Literature* 40(2):436-482.

Glewwe, Paul, and Hanan Jacoby. 1995. An economic analysis of delayed primary school enrollment and childhood malnutrition in a low income country. *Review of Economics and Statistics* 77:156-169.

Glewwe, Paul, Hanan Jacoby, and Elizabeth King. 2001. Early childhood nutrition and academic achievement: A longitudinal analysis. *Journal of Public Economics* 81(3): 345-368.

Glewwe, Paul, and Elizabeth King. 2001. The impact of early childhood nutritional status on cognitive development: Does the timing of malnutrition matter? *World Bank Economic Review* 15(1):81-114.

Golden, MH. 1994. Is complete catch-up possible for stunted malnourished children? *European Journal of Clinical Nutrition* 48(1S): S58-70.

Gramm, M. 2003. The case for regulatory rent-seeking: CRA based protests of bank mergers and acquisitions. *Public Choice* 116:367–379.

Greene, W.H. 2007. *Econometric analysis*. 6th ed. Upper Saddle River, NJ: Prentice Hall.

Guttman, L. 1954. Some necessary conditions for common factor analysis. *Psychometrika* 19:149-162.

Haddad, L., and J. Hoddinott. 1994. Women's income and boy–girl anthropometric status in the Côte d'Ivoire. *World Development* 22:543–553.

Hamermesh, Daniel S., and Jeff E. Biddle. 1994. Beauty and the labor market. *American Economic Review* 84(5):1174–1194.

Hauser, Robert M., and Hsiang-Hui Daphne Kuo. 1998. Does the gender composition of the sibships affect women's educational attainment? *The Journal of Human Resources* 33(3):644-657.

Hausman, J. 1975. An instrumental variable approach to full information estimators for linear and certain nonlinear econometric models. *Econometrica*, 43(4): 727-738.

Haveman, Robert, and Barbara Wolfe. 1995. The determinants of children attainments: A review of methods and findings. *Journal of Economic Literature* 33(4):1829-1878.

Hayton, James C., David G. Allen, and Vida Scarpello. 2004. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods* 7(2):191-205.

Heckman, James. 2008. Schools, skills, and synapses. *Economic Inquiry* 46(3): 289-324.

Heckman, James. J., and Paul A. LaFontaine . 2010. The American high school graduation rate: Trends and levels. *The Review of Economics and Statistics* 92(2): 244–262.

Hoddinott, John, and Bill Kinsey. 2001. Child growth in the time of drought. *Oxford Bulletin of Economics and Statistics* 63(4):409-436.

Hofferth, Sandra L., Lori Reid, and Frank L. Mott. 2001. The effects of early childbearing on schooling over time. *Family Planning Perspectives*33(6):259-267.

Hoffman, S. D., E. M. Foster, and F. F. Furstenberg. 1993. Re-evaluating the costs of teenage childbearing. *Demography* 30(1):1-13.

Hotz, V. J., S. McElroy, and S. Sanders. 2005. Teenage childbearing and its life cycle consequences: Exploiting a natural experiment. *Journal of Human Resources* 40(3):683-715.

Iyigun, M., and Randall P. Walsh. 2007. Endogenous gender power, household labor supply and the demographic transition. *Journal of Development Economics* 82(1):138–55.

Jansen, Kurt, M. Harris, and M. Penrose. 1987. *The Ethiopian famine*. London: Zed Press.

Kaestner, Robert. 1997. Are brothers really better? Sibling sex composition and educational attainment revisited. *The Journal of Human Resources* 32(2):250-284.

Kanbur, R., and L. Haddad. 1994. Are better off households more equal or less equal? *Oxford Economic Papers* 46(3):445–58.

Keane, Michael P., and Kenneth I. Wolpin. 2001. The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review* 42:1051–1103.

Kenkel, Donald S. 1991. Health behavior, health knowledge, and schooling. *The Journal of Political Economy* 99(2):287-305.

Kishor, S. 1999. Women's empowerment and contraceptive use in Egypt. Paper presented at the Annual Meetings of the Population Association of America, March, New York.

_________. 2000. Empowerment of women in Egypt and links to the survival and health of their infants. In *Women's empowerment and demographic processes*, ed. H. Presser and G. Sen. Oxford: Oxford University Press.

Klepinger, Daniel, Shelley Lundberg, and Robert Plotnick. 1999. How does adolescent fertility affect the human capital and wages of young women? *Journal of Human Resources* 34:421-48.

Koenen, C. Karestan , Alisa Lincoln and Allison Appleton.2006. Women's status and child well-being: A state-level analysis. *Social Science and Medicine* 63:2999-3012.

Korn, Edward L., and Barry I. Graubard. 1995. Examples of differing weighted and unweighted estimates from a sample survey. *The American Statistician* 49(3):291-295.

Kruger, Diana I. 2007. Coffee production effects on child labor and schooling in rural Brazil. *Journal of Development Economics* 82 (2): 448-463.

Lancaster, G., P. Maitra, and R. Ray. 2006. Endogenous intra-household balance of power and its impact on expenditure patterns: Evidence from India. *Economica* 73(291):435–60.

L'Engle, Kelly Ladin, Christine Jackson, and Jane D. Brown. 2006. Early adolescents' cognitive susceptibility to initiating sexual intercourse. *Perspectives on Sexual and Reproductive Health* 38(2):97-105.

Levy, Dan, and Greg Duncan. 2000. Using sibling samples to assess the effect of family childhood income on completed schooling. Working Paper no.168 (April). Evanston, Ill.: Northwestern Univ., Joint Center Policy Res.

Lu M., McGuire T.G. 2002. The productivity of outpatient treatment for substance abuse. *Journal of Human Resources* 37:309–335.,

Lundberg, S., R. Pollak, and T. Wales. 1997. Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit. *Journal of Human Resources* 32 (3):463–480.

Maccini, S., and D. Yang. 2009. Under the weather: Health, schooling and economic consequences of early-life rainfall. *American Economic Review* 99(3):1006–1026.

Maitra, Pushkar, and Ranjan Ray. 2002. The joint estimation of child participation in schooling and employment: Comparative evidence from three countries. *Oxford Development Studies* 30(1):41-62.

Manser, M., and M. Brown. 1980. Marriage and household decision making: A bargaining analysis. *International Economic Review* 21(1):31-44.

Martorell, Reynaldo.1999. The nature of child malnutrition and its long-term implications. *Food and Nutrition Bulletin* 19:288-292.

____________. 1997. Under nutrition during pregnancy and early childhood and its consequences for cognitive and behavioral development. In *Early childhood development: Investing in our children's future,* ed. M.E. Young. Elsevier: Amsterdam.

Martorell, Reynaldo, Dirk G. Schroeder, Juan A. Rivera, and Haley J. Kaplowitz. 1995. Patterns of linear growth in poor Guatemalan adolescents and children. *The Journal of Nutrition* 125(4S):1060S-1067S.

Mayer, Susan E. 1997. *What money can't buy: Family income and children's life chances*. Cambridge, Mass.: Harvard Univ. Press.

McElroy, M.1990. The empirical content of the household-bargained household behavior. *Journal of Human Resources* 25:559-583.

McElroy, M., and M. Horney. 1981. Nash-bargained household decisions, toward a generalization of the theory of demand. *International Economic Review* 22:333-349.

Meer, J., and H.S. Rosen. 2004. Insurance and the utilization of medical services. *Social Science in Medicine* 58:1623–1632.

Miao, Jing, and Walt Haney. 2004. High school graduation rates: Alternative methods and implications. *Education Policy Analysis Archives* 12:1-68.

Mocan, Naci, and Erdal Tekin. 2009. Ugly criminals. *The Review of Economics and Statistics* 92(1):15-30.

Morduch, Jonathan. 1995. Income smoothing and consumption smoothing. *Journal of Economic Perspectives* 9:103-114.

___________. 1999. Between the market and state: Can informal insurance patch the safety net? *World Bank Research Observer* 14:187-207.

___________. 2000. Sibling rivalry in Africa. *The American Economic Review, Papers and proceedings of the one hundred twelfth annual meeting of the American Economic Association* 90(2):405-409.

Mott, F. L., and W. Marsiglio. 1985. Early childbearing and completion of high school. *Family Planning Perspectives* 17:234-7.

Paris, William L., and Robert J. Willis. 1993. Daughters, education, and family budgets: Taiwan experiences. *The Journal of Human Resources* 28:863-898.

Patrinos, A. Harry, and G. Psacharopoulos. 1997. Family size, schooling and child labor in Peru-an empirical analysis. *Journal of Population Economics* 10: 289-294.

Plug, Eric. 2004. Estimating the effect of mother's schooling on children's schooling using a sample of adoptees. *The American Economic Review* 94(1):358-368.

Porter, Catherine[49]. 2007. Long-term consequences of severe shocks in childhood: Evidence from the Ethiopian famine of 1984. Center for the Study of African Economies, Economics Department, Oxford University.

Quisumbing, A., and J. Maluccio. 2003. Resources at marriage and intra-household allocation: Evidence from Bangladesh, Ethiopia, Indonesia, and South Africa. *Oxford Bulletin of Economies and Statistics* 65:283–328.

Rabe-Hesketh, S., and A. Skrondal. 2008. *Multilevel and longitudinal modeling using Stata*. (Second Ed.), Stata Press.

Rabe-Hesketh, S., A. Skrondal, and A. Pickles. 2004a. *GLLAMM manual*. U.C. Berkeley Division of Bio-statistics Working Paper Series, Working Paper 160.

____________. 2004b. *Interdisciplinary statistics, generalized latent variable modeling, multi-level longitudinal and structural equation models*. Chapman &HALL/CRC.

____________.2004c. Generalized multi-level structural equation modeling. *Psychometrica* 69(2):167-190.

____________.2003. Multi-level logistic regression for polytomous data and rankings. *Psychometrica* 68(2):267-287.

Rahmato, Dessalegn. 1988. *Peasant survival strategies.* Geneva: International Institute for Relief and Development/Food for Hungry International.

Ravallion, Martin, and Quentin Wodon. 2000. Does child labor displace schooling? evidence on behavioral responses to an enrollment subsidy. *Economic Journal* 110(462):C158–C175.

Rector, Robert E., and Kirk A. Johnson. 2005. Teenage sexual abstinence and academic achievement. Conference paper available at http://www.heritage.org/Research/Welfare/whitepaper10272005-1.cfm, accessed March, 2008.

Register, Charles A., Donald R. Williams, and Paul W. Grimes. 2001. Adolescent drug use and educational attainment. *Education Economics* 9(1):1-18.

Rencher, A. C. 1995. *Methods of multivariate analysis*. New York, Wiley.

Rosenzweig, Mark. R., and Kenneth I. Wolpin. 1980. Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica* 48(1):227-240.

[49] The author requested that the document should not yet be quoted.

Sabia, Joseph J. 2007a. Early adolescent sex and diminished school attachment: Selection or spillovers? *Southern Economic Journal* 74(1):239–268.

____________. 2007b. Reading, writing, and sex: The effect of losing virginity on adolescent academic performance. *Economic Inquiry* 45(4):647-670.

Sargan J. D. 1958. Estimating economic relationships using instrumental variables. *Econometrica* 26:393-415.

Savage, E., and D.J. Wright. 2003. Moral hazard and adverse selection in Australian private hospitals: 1989–1990. *Journal of Health Economics* 22:331–359.

Schultz, T. 1990. Testing the neoclassical model of family labor supply and fertility. *Journal of Human Resources* 25(4):599–634.

Schvaneveldt, Paul L., Brent C. Miller, E. Helen Berry, and Thomas R. Lee. 2001. Academic goals, achievement, and age at first sexual intercourse: Longitudinal, bidirectional influences. *Adolescence* 36(144):767-87.

Shafik, M. Najeeb. 2007. Household schooling and child labor decisions in rural Bangladesh. *Journal of Asian Economics* 18:946-966.

Shea, D., J. Terza, B. Stuart, and B. Briesacher. 2007. Estimating the effects of prescription drug. Coverage for Medicare beneficiaries. *Health Services Research* 43:933–949.

Shin, J., and S. Moon. 2007. Do HMO plans reduce health care expenditure in the private sector. *Economic Inquiry* 45:82–99.

Smith, L., and L. Haddad. 2000. Explaining child malnutrition in developing countries: A cross-country analysis. Research Report 111. Washington, D.C.: International Food Policy Research Institute.

Smith, Lisa C., Usha Ramakrishnana, Aida Ndiaye, Lawrence Haddad, and Reynaldo Martorell. 2003. The importance of women's status for child nutrition in developing countries. Research Report 131. Washington, D.C.: International Food Policy Research Institute.

Straus, J., and D. Thomas. 1995. Human resources: Empirical modeling of household and family decisions. In *Handbook of Development Economics* Vol. 3, ed. J.R. Behrman and T.N. Srinivasan. Amsterdam: North Holland.

Stuart, Bruce C., Jalpa A. Doshi, and Joseph V. Terza .2009. Assessing the impact of drug use on hospital costs. *Health Services Research* 44(1): 128-144.

Terza, V. Joseph, Anirban Basu, and Paul J. Rathouz. 2008. Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics* 27(3):531-543.

Thomas, D. 1990. Intra-household allocation: An inferential approach. *Journal of Human Resources* 25(4):635–64.

_________. 1994. Like father, like son; like mother, like daughter: Parental resources and child height. *Journal of Human Resources* 29(4):950–88.

_________.1997. Incomes, expenditures, and health outcomes: Evidence on intra-household resource allocation. In *Intra-household resource allocation: Methods, models, and policy*, ed. L. Haddad, J. Hoddinott, and H. Alderman. Baltimore, Md., U.S.A.: Johns Hopkins University Press for the International Food Policy Research Institute.

Townsend, R. 1995. Consumption Insurance: An evaluation of risk bearing systems in low income countries. *Journal of Economic Perspectives* 9:83-102.

Ulph, D. 1988. A general non-cooperative Nash model of household consumption behavior. Mimeo, University of Bristol, UK.

Varadharajan, Sowmya. 2003. The pitfalls of bargaining power in intra-household analysis. Mimeo, Department of Economics, Cornell University.

Velicer, W. F., and D.N. Jackson. 1990. Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research* 25(1):1-28.

Webb, Patrick, von Joachim Braun, and Yisehac Yohannes. 1992. Famine in Ethiopia: Policy implications of coping failure at the national and household level. International Food Policy Research Institute, Research Report 92.

Wooldridge, J.M. 1997. On two stage least squares estimation of the average treatment effect in a random coefficient model. *Economics Letters* 56:129–133.

_____________. 2002. *Econometric analysis of cross section and panel data*. Cambridge Massachusetts: The MIT Press

_____________. 2003. Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Economics Letters* 79:185-191.

VITA

Solomon Tesfay Tesfu was born in Ethiopia. He earned his Bachelors degree in Economics from Addis Ababa University in 1996. After working for the Secretariat of the Ethiopian Parliament for one year, Solomon returned to Addis Ababa University with a graduate fellowship from the African Economic Research Consortium to earn an M.Sc in Economics in 1999. Upon graduation he joined the department of Economics at Addis Ababa University as a lecturer where he taught several undergraduate courses. Before he left to do his PhD in the fall of 2005, Solomon was serving as elected chairman of the Department of Economics.

While pursuing his graduate studies in the Department of Economics of Georgia State University, Solomon worked as a graduate research and teaching assistant where he also taught principles of microeconomics class. As a graduate student he won the 'Harold Ball Award for Distinguished Performance in Quantitative Economics' in 2008. In recognition of his outstanding academic performance, Solomon has also been awarded the Carolyn McClain Young Leadership Fund Award from the Andrew Young School of Policy Studies for 4 years for the first time since the award was introduced.

Solomon's past involvement in research activities includes coordinating various household and firm surveys in Ethiopia for international and local organizations such as the International Food Policy Research Institute, the World Bank and the Confederation of the Ethiopian Trade Unions. His areas of research interest include Microeconometrics, Development Economics, Labor Economics, Public Finance and Experimental Economics. He can be reached at stesfu1@gmail.com.